Building Teacher Teams: Evidence of Positive Spillovers from More Effective Colleagues

Min Sun
University of Washington

Susanna Loeb
Stanford University

Jason Grissom
Vanderbilt University

Acknowledgement

Min Sun's contribution to this study is supported by a grant from the National Science Foundation under Grant No. DRL-1506494. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

In addition, we would like to acknowledge very helpful comments and feedback from seminar participants of the Center for Education Policy Analysis at Stanford University and Labor and Development Economics at University of Washington, Seattle. We also thank Ilana Horn, Mimi Engel, Kenneth Frank, and Jing Liu for helpful conversations about theoretical framing and data analysis.

Building Teacher Teams: Evidence of Positive Spillovers from More Effective Colleagues

Abstract

Student peer effects are well documented. We know far less, however, about peer effects among teachers. We hypothesize that a relatively effective teacher may positively affect the performance of their peers, while a relatively ineffective teacher may negatively impact the performance of other teachers with whom they work closely. Utilizing a decade of data on teacher transfers between schools that result in changes of peers when transfer teachers enter grade-level team in the new school, we find evidence of strong positive spillover effects associated with the introduction of peers who are more effective than the incumbent teacher himself or herself. However, the incumbent teacher's students are not meaningfully disadvantaged by the entry of relatively ineffective peers. This finding provides initial evidence that mixing teachers with diverse performance levels can be a strategy for increasing student achievement in the aggregate. These results are robust to several student sorting and teacher selection issues.

Keywords: Teacher Spillovers, Peer Effects, Teacher Transfer, Teacher Quality

Research provides persuasive evidence on teachers' contributions to student achievement (Aaronson, Barrow, & Sander, 2007; Koretz, 2002; McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004; McCaffrey, Sass, & Lockwood, 2009; Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004; Sanders & Rivers, 1996). Yet alongside a large research base showing evidence of peer effects in other workplaces (see Herbst & Mas, 2015), an emerging body of research suggests that student achievement is a function not just of one's own classroom teacher but of the combined effort of the classroom teacher and others with whom he or she works. The quality of a teacher's colleagues, for example, is correlated with the test score gains made by that teacher's students (Jackson & Bruegmann, 2009). Teachers' instructional expertise diffuses through professional interactions, and thus can change colleagues' classroom practices (Author, 2013). Teachers' collaboration with one another within teams can increase their effectiveness as measured by raising student achievement gains (Author, 2015).

Despite this initial evidence, our understanding of teacher spillover effects remains limited. If spillover effects are non-negligible, ignoring them means underestimating the impact of effective teachers. By focusing only on effects on students in a teacher's own classroom, evaluations of efforts to increase teacher quality or its equitable distribution across schools, such as the U.S. Department of Education's Talent Transfer Initiatives (Glazerman, Protik, Teh, Bruch, & Max, 2013), may underestimate those efforts' total impacts. Assuming no spillover effects in models that states and districts use to measure teachers' "value added" to student achievement may not be appropriate (Jackson & Bruegmann, 2009; Yuan, 2015). Additionally, failure to recognize spillovers among teachers on grade-level or subject-area teams within a school may lead school leaders to miss an important opportunity to strategically build teacher teams in ways that augment all students' learning.

This study examines teacher spillover effects using longitudinal administrative data from Miami-Dade County Public Schools (M-DCPS). We apply insights from the economic literature on employee peer effects in other workplaces, which emphasizes the roles of social pressure and knowledge spillover as means for employees to affect the productivity of one another, to model spillovers in the context of teacher work. We then test these models using the case of teacher transfers from other schools in M-DCPS onto existing grade-level teams of existing teachers in elementary and middle schools. The idea behind this test is that if the presence of a more effective teacher on one's team impacts other teachers' own performance, the arrival of a new peer provides an opportunity to observe evidence of this spillover. We ask: Does the effectiveness of a new transfer teacher spill over into other classrooms in that grade? More specifically, does a transfer teacher's entry into a grade-level team affect the achievement of students of incumbent teachers (i.e., those already in the school), and how do these effects depend on the relative effectiveness of transfer and incumbent teachers?

We examine four different potential types of spillovers. First, we look at the average spillover effects of new transfer teachers. This "linear-in-means" model assumes that with the arrival of an effective peer, all incumbent teachers will improve, and conversely, the arrival of an ineffective peer will hurt all others' outcomes. We then consider the non-linearity of spillover effects depending on the difference in prior stable effectiveness between new transfers and incumbent teachers—the "relatively effective" and "relatively ineffective" models. The "relatively effective" approach models how incumbent teachers' effectiveness changes in relationship to the degree to which the new peer is more effective than they are. This model could reflect knowledge transfer from more effective to less effective teachers. Similarly, the "relatively ineffective" approach measures the effect of the degree to which the new transfer is less effective. This model could capture a drain on incumbent teachers from having less effective teachers enter their grade. In contrast to the relative approach, we lastly examine the variation of spillover effects depending on the absolute effectiveness of focal (incumbent) teachers. We use “focal” teachers interchangeably with “incumbent” teachers hereafter to refer to those who are already in the grade when the new transfer joins the team and whose students’ achievement gains are the outcome measures of the analysis. This “absolute effectiveness” model evaluates which types of teachers are more or less responsive to peers’ effectiveness. Less effective teachers may be more affected by the performance of new teachers, because they need greater support from their peers or are more easily influenced.

Although we find some evidence of positive “linear-in-means” effects, we find stronger evidence of positive spillover effects associated with the introduction of relatively effective peers into a teacher group. If a student has a new peer teacher at the same grade level who is about one standard deviation more effective than that of his or her own teacher, this student would have a 1.9 or 2.8 percent of a standard deviation increase in math test scores. This spillover effect is about 23% or 29% of the student’s own teacher’s effect on his or her achievement gains. We also find that effects are asymmetrical; although teachers benefit from a relatively effective peer, their students are not meaningfully disadvantaged by the presence of relatively ineffective peer. This finding implies the way of grouping teachers to maximize all students’ learning is to mix teachers with diverse performance. In keeping with the importance of relatively effective peers, we also find some evidence that low-performing teachers are more responsive to the composition of his or her peer colleagues than high-performing teachers. Having an effective peer teacher particularly benefits students assigned to low-performing teachers.

In what follows, we first review the literature on spillover effects among employees in schools and other workplaces. Next, we describe the four types of spillover in more detail, motivated by possible spillover mechanisms among teachers. We then describe the data and analytic strategies for testing these models. Lastly, we discuss the main findings and their policy and research implications.

Spillover Effects among Employees in Schools and Other Workplaces

A large body of research finds evidence of peer effects on worker productivity in both high-skilled and low-skilled occupations and in a variety of experimental contexts (e.g., Battu, Belfield, & Sloane, 2003; Bauer & Vorell, 2010; De Grip & Sauermann, 2012; Herbst & Mas, 2015; Kurada & Yamamoto, 2013; Stoyanov & Zubanov, 2012). Explanations for peer effects in the workplace center on two mechanisms: Social pressure and knowledge transfer (Author, 2014; Cornelissen, Dustmann, & Schönberg, 2013). Social motivation works either by providing relatively low-performing workers with incentives to work more to keep up with other coworkers, or by making high-performing workers reduce their efforts to conform to the group norm. Knowledge transfer, on the other hand, is a process in which workers learn job-relevant knowledge or skills that make them more productive from observing or interacting with coworkers. Research findings are consistent with both mechanisms. For example, as evidence of social pressure, Falk and Ichino's (2006) study of short-term workers stuffing envelopes shows that the presence of a more productive peer working nearby compels less productive workers to work more quickly. Similarly, Mas and Moretti's (2009) study of supermarket cashiers finds that introducing a faster cashier into a shift increases the pace of scanning among others on the shift. These gains are limited to workers in the productive cashier's line-of-sight—suggesting the increase in peer productivity comes from a kind of monitoring pressure—and are concentrated among those peers with whom he or she works more frequently. Kurada and Yamamoto (2013) provide another example of conforming to the group norm. When employees were transferred from Japan to European branches of the same global firms, these employees significantly reduce their work hours, resulting from behavioral influences of locally hired staff. The reduction in hours highly depends on the level of the interactions between the transfers and local peers. Substantial research also shows evidence of spillovers consistent with a knowledge transfer mechanism, including studies of the transmission of knowledge learned during a formal training program to other employees (De Grip & Sauermann, 2012) and persistent gains to the productivity of Danish manufacturing firms from hiring high-skilled employees from more productive firms (Stoyanov & Zubanov, 2012).

We know far less about peer effects among teachers in schools, though the collaborative nature of teaching and its substantial on-the-job learning component make teaching a conducive context for spillovers among coworkers (Cornelissen, Dustmann, & Schönberg, 2013). Indeed, using longitudinal elementary school teacher and student data, Jackson and Bruegmann (2009) find that students have larger test score gains when their teachers have more effective colleagues, with the historical peer quality (i.e., estimated value-added from an out-of-sample pre-period) for less experienced teachers explaining about 20 percent of students' own-teacher effects. While their study suggests that peer learning is the major venue for the transmission of the peer effects that they observe, it does not directly measure the knowledge diffusion or peer learning, or the potential heterogeneity of spillover under different peer compositions.

A related strand of quantitative research uses teacher network data to provide more direct evidence on the diffusion of instructional expertise among teachers. Author (2013) identifies that spillover effects of professional development programs through teacher collaboration can be as large as the program direct effects on changing participating teachers' classroom instruction. Using data from an experimental study of a writing professional development program in 39 schools, the study shows that exposure to colleagues' expertise gained from prior-year professional development significantly increases the breadth of the writing purposes taught by a teacher and the diversity of active learning strategies to engage students in the writing processes. Moreover, teachers whose prior implementation of a new intervention was far from the desired practices responded more to direct participation in organized professional development, while teachers whose prior implementation was more advanced responded more to the sharing of promising practices and engaging in in-depth discussion with colleagues (Author, 2012). These findings provide initial evidence on the heterogeneity of peer influences depending on the level of focal teachers' prior teaching practices; however, they do not include student learning outcome measures; thus, it is unclear if the changes in teachers' self-reported instructional practices can later be transformed into changes in student outcomes.

Not all studies of peers find positive effects. For instance, a study of clustering among Teach for America's (TFA) corps members in disadvantaged schools finds no spillover effects on performance resulting from this placement strategy (Hansen, Backes, Brady, & Xu, 2014). However, in sum, empirical studies of worker peer effects are motivated by examining employees' contribution to organizational productivity beyond individual knowledge, personal attributes and behaviors that produce economic and social values. Rather, these studies highlight that individuals' human capital can have direct effects on their own productivity as well as may increase the productivity of other employees. As a result, organizational decisions about grouping employees into teams can have implications for overall productivity.

Modeling Teacher Spillover Effects

Spillover as a mechanism of increasing organizational productivity calls for an in-depth investigation of the structure of peers and the heterogeneous effects of teacher peers in schools. We hypothesize four types of peer influences, motivated by the research on social pressure and knowledge transfer mechanisms (e.g., Author, 2014; Koedel, 2009; Jackson & Bruegmann, 2009; Mas & Moretti, 2009).

First, the most common approach to modeling peer effects is the "linear-in-means" model (e.g., Graham, 2008; Sacerdote, 2001; Summers & Wolfe, 1977), which hypothesizes that an individual's outcomes are a function of the average outcomes and characteristics of his or her peers. In a teacher grade-level team, the linear-in-means approach implies that with the arrival of an effective peer, all incumbent teachers in the team will improve their outcomes, and conversely, the arrival of an ineffective peer will hurt all others' outcomes (Hoxby & Weingarth, 2005). This model is consistent with a mechanism of joint production in which various tasks that would promote student learning are distributed across teachers. For example, teachers may co-teach, co-plan, and share duties outside of their own classrooms (e.g., organizing math club). In many schools, teachers also work together to develop curriculum materials and analyze students' assessment data (Author, 2015). The addition of an effective peer could increase the overall productivity of joint activities, while adding a worse peer could reduce this collective productivity. However, while there is some theoretical defense for this linear-in-means model, it fails to capture the knowledge transfer or social pressure mechanism that are at the heart of research on worker peer effects since it doesn't measure how strong the peer is relative to the focal worker.

Peer effects may be nonlinear, varying for different individuals (Carrell et al., 2009; Hoxby & Weingarth, 2005; Imberman, Kugler, & Sacerdote, 2012; Mas & Moretti, 2009). We examine whether peer teachers may have different effects on their grade-level team colleagues depending on their “relative effectiveness”. What may matter for whether there is spillover is how much more or less effective the new colleague is than the teacher already in the team.

As our second model, we consider the case that a “relatively effective” new peer—that is, a teacher transferring into the grade-level team with higher teaching effectiveness than a focal incumbent teacher—could affect the achievement of the focal teacher’s students. Although the introduction of a relatively effective peer could worsen outcomes when focal teachers engage in “invidious comparisons” that undermine their confidence or sense of efficacy and, in turn, their effort level (Hoxby & Weingarth, 2005), several other potential mechanisms make benefits from working with a more effective peer more likely. One such mechanism that is likely to be especially important in teacher grade-level teams is knowledge transfer or peer learning (Jackson & Bruegmann, 2009). Working with other teachers in teams or professional learning communities provides opportunities for information about effective instructional practices to be disseminated from one teacher to another (Author, 2013; Author, 2015). Working together in teams allows teachers to share curricular materials, to discuss strategies for instruction or classroom management, or to model teaching practices for one another (Coburn & Russell, 2008). Each of these transfer mechanisms would benefit less effective teachers with the opportunity to work with a relatively effective colleague. Moreover, the presence of a relatively effective colleague may increase a less effective teacher’s motivation to work harder or seek out new strategies or techniques to increase his or her own effectiveness, through either friendly competition with the colleagues or being influenced by this colleague’s enthusiasm for teaching. The social pressure of not wanting to be perceived by colleagues as less productive or uncooperative may motivate less effective teacher to improve when a relatively effective colleague is present (Mas & Moretti, 2009).

Third, we consider the case that the entrance of a relatively ineffective new peer could affect a teacher's students. While it is possible that the arrival of a less effective peer could increase a colleague's performance by motivating this colleague to work harder to compensate for the lower productivity of the new peer, this ineffective peer could have negative impacts. Knowledge transfer is asymmetric. Although there is always something that one teacher can learn from the other, knowledge typically flows from more knowledgeable or productive individuals to those who are less so (Author, 2013; Conley & Udry, 2010). An ineffective peer is thus less likely to provide the more effective colleague with productivity-enhancing insights. However, this peer may impose costs on his or her more effective colleagues by taking up their time or attention in attempting to learn from them. At the same time, ineffective peers are less likely to affect their colleagues positively via prosocial pressure because they do not provide positive reference point that motivates other teachers to emulate them.

Fourth and finally, we examine how peer effects vary depending on the "absolute effectiveness" of *focal* teachers. Although there are incentives for less productive incumbent teachers to "free-ride," easing the pace when a productive peer comes in (Mas & Moretti, 2009), we hypothesize that many less effective teachers will work to minimize productivity differentials with their more effective peers, because they take pride in the social good of their profession. Motivated by prosocial pressure and having more opportunities to receive knowledge as described previously, we anticipate the less effective incumbent teachers, on average, are more likely to accept the positive influence from the new peers. In contrast, effective incumbent teachers are less affected by their peers because they may be less motivated to turn to their peers for supports or fewer opportunities to receive constructive help. These dynamics would lead to variation in spillover effects by the absolute effectiveness of the focal teachers.

Data and Sample

Our data come from M-DCPS, the fourth-largest school district in the United States, and cover the school years from 2003-04 through 2012-13. We focus on math teachers in grades 3–8 who can be linked to students for whom we have state standardized test scores in math. The data cover about 1.15 million student-year observations over the 10 years.

Our analysis focuses on estimation of spillover effects in math for several reasons. First, prior studies show that teachers generally have a stronger effect on math achievement than on reading (e.g., Nye, Konstantopoulos, & Hedges, 2004). Analysis of the data used in this study similarly suggest that the estimated effect of a student's classroom teacher on test scores is only about one-third to one-half as large in reading as in math. Second, mathematics teaching may provide a context more conducive to spillover effects than teaching in other subjects. Research has documented the distributed nature of math teaching in many schools, with teachers working together to set goals, choose instructional activities, design assessment instruments, and interpret evidence of learning (Cobb, de Silva Lamberg, & Dean, 2003). More so than many other subjects, there is widespread agreement on appropriate content, sequence, and pedagogy, which means both greater opportunities to coordinate across classrooms and greater likelihood that teachers are following similar curricula and routines. This consensus can provide a helpful basis for peer learning (McLaughlin & Talbert, 2001; Siskin, 1994). Math teachers frequently have conversations with one another about teaching and students' interactions with material across classrooms that provide opportunities to learn from one another (Horn, 2005; Horn & Little, 2010). These collegial interactions have been linked to improvement in math teachers' instructional practice (Author, 2014). Moreover, at least one prior study shows much greater impacts of collaboration with colleagues around instruction on effectiveness in math than in reading (Author, 2015, pp.28–30). While we focus on math in the main text for the purpose of brevity, we provide details of our analyses of reading achievement in Appendix A. Although overall our findings are weaker in reading than in math, we do find similar effects in the final "absolute effectiveness" model.

Table 1 describes the sample. Approximately nine percent of students are white; 25 percent black; 65 percent Hispanic; 49 percent female; 13 percent with limited English proficiency; 72 percent eligible for subsidized lunch (FRPL); and 12 percent with special education needs. Besides conventional elementary and middle schools, M-DCPS has K-8 schools and combination schools (middle and high schools). Across these different school types, a math teacher in elementary grades (3-5) typically works with one group of students across multiple subjects, while a math teacher in secondary grades (6-8) typically works with multiple groups of students within one subject area. Close to 60 percent of students are enrolled in elementary grades, with the rest in middle grades. We define a teacher's primary grade level as the grade for which she teaches the largest number of students in a given year.[1]

[TABLE 1 HERE]

We measure teachers' annual performance in raising students' math test scores using value-added scores. Our preferred value-added model estimates teacher-by-year fixed effects, accounting for students' test scores in math and reading in the prior year, demographics, English proficiency, and disability status, as well as the averages of these variables at both classroom and school levels (see Appendix B). This model adjusts teacher effect estimates for nonrandom assignment to students based on students' time-varying and invariant characteristics and school contexts. This model outperforms other popular value-added models and the student percentile growth model when nonrandom assignment of students exists (Chetty, Friedman, & Rockoff, 2014; Guarino, Reckase, Stacy, & Wooldridge, 2015). To further confirm that our spillover estimates do not vary depending on value-added models, we construct alternative value-added models with either student or school fixed effects. These alternative models yield similar estimates of spillover effects to our preferred model.[2]

After obtaining the teacher-by-year fixed effect estimates, we then shrink the estimates using the empirical Bayes (EB) methods to adjust for sampling and measurement errors and bring imprecise estimates closer to the mean (see Author, 2012, for a description of the shrinking method). After shrinking the value-added estimates, we standardize them to have a mean of 0 and a standard deviation of 1 in each year to facilitate interpretation. We acknowledge that EB estimators do not always reduce the misclassification of teachers, particularly under nonrandom teacher assignment (Guarino, Maxfield, Reckase, Thompson, & Wooldridge, 2015). In our study, the Spearman's rank correlation coefficient between the shrinkage estimates and teacher fixed effect estimates without shrinkage is 0.998. The strong correlation between shrinkage and fixed effect estimates is because we exclude teachers whose class size was less than 10 students per year. Unsurprisingly, when we replace the EB estimates with teacher fixed effects in Equation 1-3, results (as included in the Online Appendix Table OA-1) are very consistent with main results using shrinkage estimators.

We then average three lagged value-added measures to account for concerns about year-to-year fluctuation of value-added measures due to the variation in true teacher performance over time and measurement error (Author, 2013). We name this aggregated measure as teachers' *prior stable effectiveness*. This prior stable effectiveness has at least two advantages. First, the stable effectiveness prior to the peer shock of new transfers avoids the reflection problem in peer effect estimation, which we will explain further in next section (Manski, 1993). Second, this stable measure mitigates the spurious relationship between new transfers and incumbent teachers due to contemporaneous shocks to all teachers at a given point in time.

There are 1,594 teacher-year transfer observations in the data that have stable teacher effectiveness measures over these ten years. As shown in Table 1, approximately 37 percent of these transfer teachers are white, 34 percent black, and 26 percent Hispanic. The total percentage of nonwhite transfer teachers is little over 63 percent, which is about eight percent higher than staying teachers. About 77 percent of transfer teachers are female, compared to 83 percent of incumbent teachers. The average transfer teachers' working experience in this district is 7.6 years, which is about 3.5 years junior than the average teaching experience in the district. Moreover, transfer teachers are, on average, less effective than incumbent teachers (-0.21 vs. -0.05), less likely to have advanced degree (master's or higher, 43 percent vs. 45 percent), but have fewer days absent from work (5.8 vs. 6.6).

Analytic Strategies

We estimate spillover effects by leveraging the peer shock to incumbent teachers due to new teachers' transferring into a teacher group. Our main analyses focus on grade-level peers, teachers who teach the same grade in the same school and year. This peer definition allows us to use different fixed effects in modeling spillover effects of new peer teachers on incumbent teachers' student achievement. For example, we use (a) school-grade fixed effects to control for stable characteristics and practices in the given grade and school (e.g., stable peer effects among continuing teachers in a given school and grade) (b) school-year fixed effects to control for other possible school-year shocks than new peers' entry (e.g., enhanced professional development or teacher collaboration in the school in a given year), (c) year fixed effects to control year-to-year variations in district policies that may influence teacher collaboration and student achievement, and (d) grade fixed effects for grade-level differences that could affect both student achievement and teacher transfer behaviors.

Peer effects may expand beyond grade-level peers. This expansion may be particularly likely in schools with strong teacher collaborative activities. However, school-level peer estimates are subject to other yearly shocks to the schools that may coincide with the arrival of new peers and cannot be easily addressed using our data. We thus focus on grade-level peer effects. To demonstrate the possible spillover beyond the grade level, we show school-level estimates in Appendix C with a caveat of weaker identification strategies.

There are three key challenges for identifying peer effects in literature: Common influences, reflection, and selection, all of which can lead to bias in the estimate of peer effects. If we were to use peer characteristics that were contemporaneous with the focal teacher effect, we would worry about common influences, for example, students having an illness at the time of the test or teachers' co-participation in professional development programs. These common influences would affect the performance of both new transfers and incumbent teachers, and appear to be a peer effect. However, since we measure the peer characteristic—value-added—prior to when the peer and focal teacher interact, these common influence problems should not bias our peer effect estimates. The reflection problem is similar. It refers to the scenario when one individual's outcome is influenced by others in a given period of time, and influences others in the same period (Manski, 1993). Because we use the peer teacher value-added prior to when he or she met the focal teacher, reflection is not a problem in our case.

The final potential source of bias—selection— is more difficult to address. Selection may bias the peer estimates in settings where peers self-select into peer groups in a manner that is unobserved to researchers. For example, new transfers may select schools with similar peers, or principals may assign new transfers to peer groups with similar productivity. This selection could cause substantial upward bias in the estimated magnitude of peer effects (Sacerdote, 2011). By controlling for incumbent teachers' prior stable effectiveness and by comparing grades within schools within a given year, we adjust for much of this selection. We conduct falsification tests to examine other potential mechanisms of teacher selection in a later section in the paper, confirming that any resulting biases have little impact on our estimates. Moreover, we use (a) school-grade fixed effects to account for the time-invariant attractiveness of a particular grade in a school, (b) school-year fixed effects to account for time-varying attractiveness of a particular school in a given year, and the year-to-year variation in vacancies in a given school due to teacher turnover or the increase in student enrollment, and (c) time-varying lagged student achievement scores to account for the possibility of transfer teachers using such information to make their selection.

"Linear-in-means": To construct the "linear-in-means" model, we model student math test score as a function of his or her teachers' prior effectiveness and the average prior stable effectiveness of new peer teachers. In particular, we model:

$$A_{ijgst} = \alpha_0 + \alpha_1 A_{ijgst-1} + \alpha_2 A_{ijgst-1}^{other} + \boldsymbol{\gamma_1}\mathbf{X}_{ijgst} + \boldsymbol{\gamma_2}\mathbf{C}_{ijgst}$$

$$+ \beta_1 {*} \theta_{jgst-1,2,3} + \beta_2 {*} \theta_{kgst-1,2,3} + \mathbf{SG}_{gs} + \pi_t + \varepsilon_{ijgst} \qquad (1)$$

where A_{ijgst} is the math exam score of student *i*, taught by incumbent teacher *j* in grade *g*, school *s*, and year *t*. This variable does not include new transfers' own students' scores on the left-hand side of the equation but only includes students' test scores of incumbent teachers, so that we can better attribute the gain in test scores to peer effects, rather than to own teachers' contribution to student achievement gains. $A_{ijgst-1}$ indicates this student's prior year math test score and $A^{other}_{ijgst-1}$ indicates his or her prior year score in the other subject (e.g., reading). $\mathbf{X}_{ijgst}$ is a vector of student *i*'s characteristics, including poverty status, whether the student is an English language learner, the student's race, gender, age, prior suspension, and prior absence. $\mathbf{C}_{ijgst}$ is a vector of student *i*'s classmates' characteristics, such as percent of students eligible for subsidized lunch, percent of students who are English language learners, percent of Hispanic, African American, Asian, and white student, percent of female students, average age, average number of days suspended, average days absent, and the average and standard deviation of prior scores in math and reading. $\theta_{jgst-1,2,3}$ is student *i*'s own teacher *j*'s value-added scores averaged over prior three years (*t*-1, *t*-2, and *t*-3)—the focal (incumbent) teachers' prior stable effectiveness. $\theta_{kgst-1,2,3}$ is the newcomer *k*'s value-added scores averaged over three years prior to transferring into this school (*t*-1, *t*-2, and *t*-3); and β_2 captures the "linear-in-means" estimate. $\mathbf{SG}_{gs}$ is the school-grade fixed effects, and π_t is the year fixed effects. We also estimate Equation 1 with the combination of school-year, and grade fixed effects. We cluster the standard errors at the school-grade-year level. ε_{ijgst} is the error term.

"Relative effectiveness": We then examine how the peer effects vary depending on the difference in effectiveness between the transfer and focal teacher —student *i*'s own incumbent teacher *j*. We define "relative effectiveness" as how much more effective the new transfer *k* was over the preceding three years than the focal teacher. "Relative ineffectiveness" is then defined as how much less effective the new transfer *k* was than the focal teacher *j*. We estimate the effects of these two types of peers separately because we suspect differential effects of "relatively effective" and "relatively ineffective" peers.

$Relative\ effectiveness_{k,\ jgst-1,2,3} \equiv D^*(\theta_{kgst-1,2,3} - \theta_{jgst-1,2,3})$

$Relative\ ineffectiveness_{k,\ jgst-1,2,3} \equiv (1-D)^*(\theta_{kgst-1,2,3} - \theta_{jgst-1,2,3})$

where D=1 if $(\theta_{kgst-1,2,3} - \theta_{jgst-1,2,3})>0$; D=0 if $(\theta_{kgst-1,2,3} - \theta_{jgst-1,2,3})<0$.

We estimate the effects of "relatively effective" and "relatively ineffective" peers using Equation 2.

$$A_{ijgst} = \alpha_0 + \alpha_1 A_{ijgst-1} + \alpha_2 A_{ijgst-1\ reading} + \boldsymbol{\gamma_1 X}_{ijgst} + \boldsymbol{\gamma_2 C}_{ijgst} + \beta_1{}^*\theta_{jgst-1,2,3}$$
$$+\beta_2{}^*\ Relative\ Effectiveness_{k,\ jgst-1,2,3} + \beta_3{}^*\ Relative\ Ineffectiveness_{k,\ jgst-1,2,3}$$
$$+ \mathbf{SG}_{gs} + \pi_t + \varepsilon_{itgst} \tag{2}$$

"Absolute effectiveness": We then test for heterogeneous effects depending on the prior stable effectiveness of incumbent teachers using an interaction term between incumbent teacher *j*'s prior stable effectiveness and new peer *k*'s prior stable effectiveness. Equation 3 illustrates the estimation model.

$$A_{ijgst} = \alpha_0 + \alpha_1 A_{ijgst-1} + \alpha_2 A_{ijgst-1\ reading} + \boldsymbol{\gamma_1 X}_{ijgst} + \boldsymbol{\gamma_2 C}_{ijgst} + \beta_1{}^*\ \theta_{jgst-1,2,3} + \beta_2{}^*\theta_{kgst-1,2,3}$$
$$+ \beta_3{}^*\theta_{jgst-1,2,3} \times \theta_{kgst-1,2,3} + \mathbf{SG}_{gs} + \pi_t + \varepsilon_{ijgst} \tag{3}$$

where β_3 indicates the amount of additional gain in student *i*'s test score that can be attributed to a new transfer teacher *k*, with one standard-deviation increase in own teacher *j*'s prior stable effectiveness. This interaction effect identifies what types of teachers more or less benefit from a high-performing peer.

We conduct robustness and falsification tests to examine how plausible teacher selection and student sorting may bias the main estimates of grade-level spillover effects. These tests are detailed in the next section.

Results

Grade-level Spillover Patterns

For each type of spillover models, we present figures that graphically illustrate the patterns and regression estimates that formalize these patterns. Figure 1 shows a positive association of the new peers' prior stable effectiveness with focal teachers' student math achievement. This pattern is formalized in the linear-in means estimates in Table 2. The average effects of a one standard deviation change in the prior stable effectiveness of the new transfer teacher on the achievement gains of students taught by incumbent teachers in the same grade, are between one percent and two percent of a standard deviation of students' math test scores. They are positive and mostly statistically significant at either the 0.10 or 0.05 level. These linear-in-means effects are between 15 percent and 29 percent of the effects of own teachers' effects (0.01/0.068 or 0.020/0.069). These percentages are consistent with Jackson and Bruegmann's (2009) estimates of between 10 and 20 percent of the own teacher effect (p.99).

[FIGURE 1 HERE]

[TABLE 2 HERE]

The size of our linear-in-means effects is somewhat smaller than Jackson and Bruegmann's (2009) estimate of approximately four percent of a standard deviation of math test scores, likely stemming from some differences in the types of peer spillovers estimated as well as the measures of teacher effectiveness. First, the peer measure in our study averages the effectiveness of only transfer teachers and thus captures the effects from new transfer teachers to other teachers at the grade level, while Jackson and Bruegmann's (2009) study averages all grade-level peers and captures the peers among all grade-level teachers. Second, our study leverages the one-year co-working experience among teachers, while Jackson and Bruegmann's (2009) peer effects may reflect co-working experience in multiple years. Third, notably the linear-in-means effects are estimated using teachers' prior stable effectiveness, which results in the smaller point estimate than current year teacher effects. This is in similar vein of the estimate of the own-teacher effect of 0.07 in Table 2, which is smaller than Jackson and Bruegmann's (2009) estimates of own teacher effect estimates in math—approximately 0.13 standard deviations.

Figure 2 illustrates the "relative effectiveness" model. The linear fit line for cases with the x-axis< 0 is close to flat, which shows a very weak relationship between "relatively ineffective" peers and focal teachers' student math achievement. In contrast, the linear fit line for cases with the x-axis greater than 0 has a steeper, more positive slope, which indicates a much stronger positive association between "relatively effective" peers and focal teachers' student achievement Table 2 provides estimates that formalize the differential effects of having peers who are more or less effective than the focal (incumbent) teacher. If a student in the class of an incumbent teachers has a new transfer teacher at the same grade level who is one standard deviation higher in prior stable effectiveness than that of their own teacher, this student would have a 1.9 or 2.8 percent of a standard deviation increase in math test scores. This spillover effect is about 23 or 29 percent of the student's own teacher effect (0.019/0.081 or 0.028/0.095). Surprisingly, if the transfer peer teacher is about one standard deviation lower than that of own teacher, this student would not be meaningfully affected by the new teacher. The "relatively ineffective" estimate is very close to zero and not statistically different from zero. An *F*-test shows that the "relatively effective" estimate is significantly different from the "relatively ineffective" estimate (F=6.88, p< 0.001).

[FIGURE 2 HERE]

Finally, Table 2 gives the variation of spillover effects by the absolute effectiveness of incumbent teachers, as indicated by the interaction term between new transfers' prior stable effectiveness and own teachers' prior stable effectiveness. It measures whether more or less effective incumbent teachers are differentially affected by transferring teachers. The significantly negative coefficients provide evidence that with one standard deviation increase in own teachers' prior effectiveness, the spillover effect from new transfer peers would decrease about 0.6 percent or 0.8 percent of one standard deviation of student test scores. In other words, new peer teachers matter less for students whose own teachers were relatively more effective, or equivalently, that they matter more for those students whose own teachers were less effective.

Figure 3 bases on Equation 3 and plots the marginal effects of new peers on focal teachers' student achievement (i.e., the predicted spillover effects) against focal teachers' prior stable effectiveness. The figure confirms a substantial heterogeneity in how teachers respond to peers: the spillover is positive and larger for low-performing teachers, and has little effect on the high-performing ones. Notably, the estimated effects are negative for just a very small number of cases and their 95% confidence intervals all include zero, suggesting that the effectiveness of high-performing teachers is, on average, not particularly hurt by the presence of low-performing peers.

[FIGURE 3 HERE]

Overall, the main findings show that teachers who are newly transferred to a grade affect the learning of students of incumbent teachers. These effects are bigger when the new teacher is more effective than the incumbent teacher, while the new teacher who is less effective has little impact on students' learning in the incumbent teacher's classrooms. The positive spillover effects are also bigger for less effective incumbent teachers.

Robustness and Falsification Tests

Teacher Sorting

Other possible shocks to the composition of grade-level peers could affect student learning and bias our estimates of peer spillover. First, it is possible that a novice teacher who just started her/his career was employed at the same time in the same grade and school as a new transfer entered the team. Equations 1-3 would drop this novice teacher and her students from the analysis, because a novice teacher did not have prior value-added scores on the right-hand side of the equations. For the same reason, the new transfer teachers without prior stable effectiveness, although being part of the new members of the teaching team, would be dropped out of the analysis too. These other new peers, including both novice teachers and new transfers without prior stable effectiveness, could be the omitted factor that confounds the grade-level phenomenon of benefiting incumbent teachers, if the entrance of these other new teachers is correlated with the prior performance of new transfer teachers who had prior stable effectiveness. To account for the influence of other new teachers, we create a continuous variable indicating the number of other new teachers at the same grade and add that to Equations 1-3. Table 3 reports the results in the columns labeled "With Other New Teachers." The point estimates and standard errors of "linear-in-means," "relatively effective/ineffective," and "absolute effectiveness" effects are quite consistent with corresponding estimates in Table 2, as are the adjusted R-squared values.[3]

[TABLE 3 HERE]

Teachers churn within schools with some teachers moving to a new grade that they did not teach in the year before (entry) and others moving out (exit) (Author, 2014). These churning teachers could affect students in much the same way as novice teachers do. To assess the degree to which the *entry* to a given grade affects the spillover estimates, we re-estimate Equations 1-3 using only incumbent teachers who stayed in the same grade and corresponding estimates are included in the columns of "Only Same-Grade Teachers." Although "linear-in-means," "relatively effective," and "absolute effectiveness" estimates are slightly larger than corresponding estimates in Table 2, their inferences are not different in any meaningful ways. The estimate of "relatively ineffective" is negative but nonsignificant in Model 1, while Model 2 estimate is significantly negative. However, these two estimates are not statistically significantly different from each other (z=1.27, *p*=0.2), based on Cohen & Cohen's test for the differences between two regression coefficients from the same sample that accounts for the covariance between these two coefficients (1983, p.479). The significant coefficient in one model specification may not be practically meaningful; thus, we interpret the inference of "relatively ineffective" consistently with that in Table 2.

A third possible shock to a given grade in a given year that could bias our estimates is teachers' *exit*. The main concern is that an ineffective teacher's moving out of a grade and school in year t is followed by an effective new peer transferring in year t+1. The increase in student achievement might stem not from the spillover of new effective transfer in year t+1 to incumbent teachers, but rather because of the removal of an ineffective teacher from this grade and school in year t. If there is a systematic pattern that departed teachers were, on average, less effective than stayed teachers in year t, *and* a new transfer was more effective than incumbent teachers in year t+1, the significant positive effect of "relatively effective" peers could be invalidated. To address this concern, we regress the *difference* in prior effectiveness between new transfers and staying teachers in year t+1 on the *difference* in prior effectiveness between departed teachers and staying teachers in year t. The point estimate is small and not significantly different from zero (β = -0.005, s.e. = 0.089; p=0.959). Thus, we find no evidence that the arrival of a more effective new teacher to a particular grade in a school is related to the performance difference between departed and staying teachers in the prior year.

Student Sorting and Other Grade-Specific Interventions

Besides teacher sorting, it is possible that our results are confounded by dynamic student sorting (or tracking) or other related grade-level interventions that cannot be fully controlled by lagged test scores, individual characteristics, and classmates' characteristics. For example, the "relatively effective" estimate can reflect that incumbent teachers get better students and also lobby for better new peers. This particular sorting would be problematic. We conduct a falsification test by regressing a student's test score in year t-1 on his/her *future* teacher's value-added in year t, controlling for this student's teacher effect in year t-1, characteristics of this student and his/her classmates' characteristics, and school-grade, year fixed effects or school-year, grade fixed effects. If there was troubling unobservable student sorting, the coefficient of *future* teacher should be statistically significant. Although the coefficients on current teacher effect are about 0.11 and significant at the 1 percent level, the coefficient on *future* teacher's effect is only 0.002 with p-value greater than 0.8. Next, to assess whether incumbent teachers lobby for better new peers, we regress incumbent teachers' prior stable performance on transfer teachers' prior stable performance, controlling for school fixed effects. The small coefficient of -0.016 is far from statistical significance (p=0.307).[4] Moreover, this point estimate suggests a negative, rather than positive relationship, which does not support the possibility that effective incumbent teachers lobby for effective new peers.

Another possibility is that a principal might assign an effective new transfer to a poor performing grade as part of his grade-specific improvement, while this principal might implement other interventions at the same time (Jackson & Bruegmann, 2009). To test whether these possibilities would invalidate the inference of peer spillover, we regress grade-level average test scores of students of incumbent teachers in year *t-1, t-2,* and *t-3* respectively on the effectiveness of *future* new transfers in year *t*. The coefficients on future new transfers range from -0.01 to 0.003 (p = 0.5 ~ 0.9), which are far from statistical significance. Taking these falsification tests together, there is little evidence on assigning new transfers as part of student sorting, lobbying for better new peers, or grade-specific interventions on student achievement.

Other Endogeneity Problems Associated with Voluntary Teacher Transfer

One might be still concerned about the spurious relationships between teacher self-selection into the school and student achievement, which cannot be fully accounted for using school-grade and school-year fixed effects, and lagged test scores. To further circumvent the problem of teacher self-selection, we leverage a unique involuntary transfer policy utilized by M-DCPS over a three-year period.

In the 2010, 2011, and 2012 school years, M-CDPS exercised a clause in its Collective Bargaining Agreement (CBA) allowing for the transfer of teachers involuntarily within the district (Author, 2014). In the summer prior to each of school years, principals provided regional administrators with the lists of teachers they wanted to transfer out of their schools, which were then forwarded to the Instructional Staffing division in the district central office, who sought a new placement for each teacher. The placement takes into account staffing needs of the receiving schools, and, in some cases, input from regional administrators, but no input from transferred teachers themselves. In each year, transferred teachers were notified of the transfers and the new placements at the very end of the summer—in many cases not until the week before the start of school. There was no time for transferred teachers to appeal, so almost all teachers complied with their new placements. The average effectiveness of involuntary transfers can be considered plausibly exogenous.

Among 153 elementary and middle school teachers who were moved involuntarily over these three years, there were 46 math teachers with value-added. We estimate Equations 1-3 for this subsample of teachers. Results are included in Table 4 and show similar patterns of spillover effect as in our main sample that includes both voluntary and involuntary transfers. While the point estimates are not statistically significant, the small sample of involuntary transfers leads to larger standard error estimates and the loss of precision of the estimation. However, the magnitude and direction of the estimated spillover effects are very similar to the main findings in Table 2. Namely, the “linear-in-means” estimate is 0.015, the “relatively effective” estimate is 0.031, the “relatively ineffective” estimate is -0.009, and the “absolute effectiveness” estimate is -0.014, in comparison to the estimates in Table 2, which are0.02, 0.028, -0.009, and -0.008, respectively.

[TABLE 4 HERE]

Alternative Measures of Teachers’ Prior Stable Effectiveness

Another concern is that measuring prior stable effectiveness by averaging teachers’ value-added over prior three years restricts our sample to a group of teachers who are relatively experienced and thus restricts our inferences of peer effects to this peer set. This aggregated measure includes teachers who have three lagged value-added measures (about 21%), and those who only have two (27%) or one lagged value-added (52%). To examine how our estimates of spillover effects vary depending on the number of lagged value-added measures available, we use either the most recent lagged value-added or the most recent two lagged value-added in the estimation. These results, included in Online Appendix Table OA-2, do not differ in any meaningful way from those presented in the Table 2 of the main text.

An alternative to averaging prior value-added is to make use of the panel nature of our data to directly estimate time-varying prior value-added for each teacher. Given the year span from 2004 through 2013, we specify eight estimates for each teacher, one estimate for every three years (e.g., 2012-2010; 2011-2009; 2010-2008; 2009-2007; 2008-2006; 2007-2005; 2006-2004; 2005-2004). Covariates are the same as specified in Equation B1 in Appendix B. We obtain both teacher fixed effects and EB estimators, and exclude teachers with less than 10 students. We then replace the original prior stable effectiveness measure with the new measure in Equations 1–3. The inferences of the new results, reported in Online Appendix Table OA-3, are consistent with our main estimates in Table 2.

In addition, to test the degree to which the new transfer teacher's prior stable effectiveness reflects the quality of her/his previous school, we derive the measure of *the average value-added of teachers in the prior-year school (excluding the new transfer teacher herself/himself)* and include it as a control in Equations 1–3. As shown in Table OA-4, none of the corresponding coefficients are statistically significant. The "linear-in-means", "relatively effective/ineffective", and "absolute effectiveness" coefficients do not change in any meaningful ways from those estimates in Table 2. These imply that the new transfer teachers' effect does not simply reflect the unobserved quality of their prior schools.

Lastly, another proxy for peer quality could be teaching experience. For each school-year-grade cell, we compute the mean teacher experience of new transfer peers. Similar to our main approach, we construct measures of "relatively experienced"—that is, the new peers are, on average, more experienced than incumbent focal teachers— and "relatively inexperienced"—that is, the new peers are, on average, less experienced than incumbent focal teachers. Online Appendix Table OA-5 show the results in two samples: the first sample includes all teachers with valid teaching experience, and the second sample includes teachers with valid teaching experience and prior stable effectiveness measure. In both samples, teachers' experience is a very weak predictor of student achievement, as indicated by the very small coefficients of students' own teachers' experience. This is consistent with a number of prior studies of teacher experience (e.g., Chetty et al., 2014; Jackson & Bruegmann, 2009; Kane, Rockoff, & Staiger, 2008). It is not surprising, then, that most of the spillover effects are statistically nonsignificant. However, the directions of the coefficients are similar to those presented in our main texts. That is, the "linear-in-means" effects are positive. Relatively experienced peer teachers have positive spillover effects. Focal (incumbent) teachers who are more experienced are less likely to be influenced by new peer teachers, while junior focal teachers are more likely to be influenced by new peer teachers. However, we interpret these patterns with great caution because most of them are statistically nonsignificant.

Testing for Different Spillover Effects in Elementary and Middle Grades

Elementary and middle grades have different organizational structures that may influence peer formation and influence among teachers. First, elementary teachers are often assigned to work with a particular group of students, while teachers in middle grades are often responsible for multiple groups of students. Second, collaboration among elementary teachers is more common within grades, while collaboration among secondary teachers is more common across grade levels, but within subject areas. These differences in the sharing of students and collaboration structure between elementary and middle grades may suggest differential spillover effects in elementary and middle grades.

We estimate the spillover effects separately by elementary and middle grades and provide the results in Table 5. Most of the results are quite similar across grade levels and to the pooled estimates (i.e., the main effects), but there is more variation across models when examined separately by school levels. In particular, for elementary grades, the Model 1 with school-grade and year fixed effects gives estimates that are very similar to the pooled estimates, while the Model 2 with the school-year and grade fixed effects gives estimates that are generally lower in magnitude, with the exception of the "relatively ineffective" and "absolute effectiveness" estimates. For the middle grades, on the other hand, the school-year and grade fixed effects estimates tend to be larger in magnitude than those with school-grade and year fixed effects. The reason for these differences may be that elementary grades typically have fewer new transfers in a given year and thus less variation in transfer teachers' effectiveness within a school and year. Therefore, the estimates from the school-year grade model that relies on the variation within a school year consistently generate smaller estimates than the school-grade year model that relies on the variation over time. In contrast, middle school grades may have more teachers transferred in a given school and year, and therefore, the Model 2 that uses the variation in transferred teachers' effectiveness within a school and year consistently generate larger estimates.

[TABLE 5 HERE]

Discussion and Conclusion

Teaching has been described as an isolated practice with few interactions among teachers (e.g., Lortie, 1975). Yet, recent reforms have worked to increase teacher collaboration, and some recent work has demonstrated effects of teachers on each other (e.g., Author, 2013; Author, 2015; Jackson & Bruegmann, 2009). Our study investigates the effects of new transfer teachers to grade-level teams, asking specifically whether having a more effective teacher entering the grade improves the learning of students of other incumbent teachers before and after the new teacher entered. Overall, we find strong and consistent evidence of positive spillover effects, as more effective teachers boost students of other teachers in the grade.

In particular, if a student has a new peer teacher at the same grade level who is about one standard deviation more effective than his or her own teacher, this student would have a 1.9 or 2.8 percent of a standard deviation increase in math test scores. This spillover effect is about 23 percent to 29 percent of this student's own teacher effect. This positive spillover effect from relatively effective peers is robust to (a) using school-grade fixed effects to account for the time-invariant characteristics of a particular school and grade, or (b) using school-year fixed effects to account for time-varying characteristics of a particular school and yearly variations in school conditions. We also present a variety of falsification tests to show that the results are unlikely to be biased by nonrandom student sorting and by endogenous teachers' movement across grades and schools.

Although relatively effective peers have positive spillovers, students of incumbent teachers are not particularly disadvantaged by the presence of relatively ineffective teacher peers. Moreover, low-performing teachers seem more responsive to the composition of peers than high-performing teachers. With one standard deviation decrease in own teachers' prior effectiveness, the spillover effect from new transfers would increase about 0.6 percent or 0.8 percent of one standard deviation of student test scores. These findings imply that strategic grouping of teachers to potentially maximize all students' learning in aggregate is to pair ineffective teachers with more effective colleagues.

Although this study could not provide direct evidence on spillover mechanisms, the significant "relatively effective" estimates support the explanations both of knowledge spillover from effective teachers to less effective colleagues and of prosocial motivation and peer pressure. These findings are consistent with prior studies of teacher professional networks that facilitate the adoption of new instructional technology in schools (e.g., Frank, Zhao, & Borman, 2004) and promote the diffusion of reform-preferred instructional practices (Author, 2013). Moreover, the findings of the "linear-in-means" model could be interpreted as evidence of joint production, with the arrival of an effective new peer increasing the team's average productivity.

Future studies can continue to examine what drives spillover so that policymakers can better design teacher incentive and professional development programs or manipulate teacher assignments to magnify effective teachers' contribution to teacher teams' performance. For example, if knowledge spillover is the predominant mechanism, school leaders can organize effective professional development programs and professional learning communities within schools to facilitate the diffusion of instructional expertise (e.g., Author, 2013). If motivation is the primary mechanism, strategies of making effective teachers visible to their colleagues, such as recognizing effective teachers through differential pay or career ladder system, could be useful.

Besides unraveling the mechanisms of spillover effects, research is needed to investigate the long-run effect of spillovers. What would be the spillover effect beyond the first year of transfer to the new schools? Would spillover effects be augmented when teachers have longer period of time to collaborate, or spillover effects decay in the second and third years after the shock effect of new peers to the grade team disappears? Moreover, we know little about the conditions under which positive spillover effects can be magnified and sustained, such as in schools that have systemic structure to promote collaboration among teachers (e.g., coherent curriculum and common planning time).

While research on peer effects across teachers is still in its infancy, findings in this study launch a series of policy discussions on teacher talent management. Our analyses do not necessarily show that omitting spillover effects would generate biased estimates of own teachers' contribution to students' learning, however, the study does show that value-added models used by states and districts that assume no spillover effects among teachers may not be appropriate for capturing teachers' full contribution to student learning, which comes both from direct effects on teachers' own students and indirect effects on other students through their influence on peer teachers (see similar suggestions in Jackson & Bruegmann, 2009; Yuan, 2015). Moreover, the existence of spillovers casts doubt on policies that only incent teachers' contributions to their own students' learning, because this type of policies may discourage teachers from making positive impacts beyond their own classrooms. Third, the positive effects of more effective peers combined with the essentially null effects of less effective peers suggests that strategic teacher grouping could benefit all students. Finally, the existence of spillover effects highlight additional benefits of retaining most effective teachers because they pay off for the school in ways that are not accounted for by the individual-based teacher evaluation framework, particularly benefiting their least effective peer teachers.

Notes

1. If a teacher had multiple grades with the same number of students, we use the lowest grade. This applies to 1.7 percent of teachers in our sample—a very small fraction of our sample.

2. The results are available from the authors upon request.

3. We also try specifications that include an indicator for having any of other new teachers instead of the number of other new teachers and find very similar results. Detailed results are available upon request.

4. This coefficient is relatively consistent across different model specifications with different fixed effects. For example, the school-grade year combination yields an estimate of -.017 (s.e.=0.02, p=0.84), and the school-year grade combination yields an estimate of -0.048 (s.e.=0.03, p=1.58).

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics, 25*(1), 95-135.

Author. (2012).

Author. (2013).

Author. (2014).

Author. (2015).

Battu, H., Belfield, C.R., & Sloane, P.J. (2003). Human capital spillovers within the workplace: Evidence from Great Britain. *Oxford Bulletin of Economics and Statistics, 65*(5), 575-594.

Bauer, T. K., & Vorell, M. (2010). *External effects of education: Human capital spillovers in regions and firms (RUHR #195).* Bochum, Germany: Ruhr-Universität Bochum (RUB).

Carrell, S. E., Fullerton R.L., & West, J. E. (2009). Does your cohort matter? Measuring peer effects in college achievement. *Journal of Labor Economics, 27*(3), 439-464.

Chetty, R., Friedman, J.N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review, 104*(9), 2593-2632.

Cobb, P., McClain, K., de Silva Lamberg, T., & Dean, C. (2003). Situating teachers' instructional practices in the institutional setting of the school and district. *Educational Researcher*, *32*(6), 13-24.

Cohen, J., & Cohen, P.(1983). *Applied multiple regression/correlation analysis for the behavioral sciences (2nd edition).* New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

Conley, T. G., & Udry, C. R. (2010). Learning about a new technology: Pineapple in Ghana. *American Economic Review*, *100*(1), 35-69.

Cornelissen, T., Dustmann, C., & Schönberg, U. (2013). *Peer effects in the workplace (IZA Discussion Paper No. 7617).* Retrieved from http://ftp.iza.org/dp7617.pdf

De Grip, A., & Sauermann, J. (2012). The effects of training on own and co-worker productivity: Evidence from a field experiment. *The Economic Journal*, *122*(560), 376-399.

Falk, A., & Ichino, A. (2006). Clean evidence on peer effects. *Journal of Labor Economics, 24*(1), 39-57.

Glazerman, S., Protik, A., Teh, B., Bruch, J., & Max, J. (2013). *Transfer incentives for High-performing Teachers: Final results from a multisite experiment (NCEE 2014-4003*). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Graham B. S. (2008). Identifying social interactions through conditional variance restrictions. *Econometrica, 76*, 643-660.

Guarino, C. M., Maxfield, M., Reckase, M.D., Thompson, P. N., & Wooldridge, J. M. (2015). An evaluation of Empirical Bayes's estimation of value-added teacher performance measures. *Journal of Educational and Behavioral Statistics, 40*(2), 190-222.

Guarino, C., Reckase, M., Stacy, B., & Wooldridge, J.(2015). A comparison of student growth percentile and value-added models of teacher performance. *Statistics and Public Policy. 2*(1), retrieved at http://dx.doi.org/10.1080/2330443X.2015.1034820.

Hansen, M., Backes, B., Brady, V., & Xu, Z. (2014). *Examining* spillover *effects from Teach for America corps members in Miami-Dade County Public Schools. (CALDER Working Paper No.113)*. Washington, D.C.: National Center for Analysis of Longitudinal Data in Education Research, American Institute for Research.

Horn, I. S. (2005). Learning on the job: A situated account of teacher learning in high school mathematics departments. *Cognition and Instruction, 23*(2), 207-236.

Horn, I.S. & Little, J.W. (2010, March). Attending to problems of practice: Routines and resources for professional learning in teachers' workplace interactions. *American Educational Research Journal*, *47*(1), 181-217.

Hoxby, C. M., & Weingarth, G. (2005). *Taking race out of the equation: School reassignment and the structure of peer effects.* retrieved from https://www.aeaweb.org/assa/2006/0108_1300_0803.pdf.

Imberman, S. A., Kugler, A. D., & Sacerdote, B. I. (2012). Katrina's children: Evidence on the structure of peer effects from hurricane evacuees. *American Economic Review, 102*(5): 2048-2082.

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.

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*(6): 615–31.

Koedel, C. (2009). An empirical analysis of teacher spillover effects in secondary school. *Economics of Education Review*, *28*(6), 682-692.

Koretz, D. M. (2002). Limitations in the use of achievement tests as measures of educators' productivity. *Journal of Human Resources, 37*, 752-777.

Kuroda, S., & Yamamoto, I. (2013). Do peers affect determination of work hours? Evidence based on unique employee data from global Japanese firms in Europe. *Journal of Labor Research, 34*(3), 359-388

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

Manski, C.F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies, 60*(3), 531-542.

Mas, A., & Moretti, E. (2009). Peers at work. *American Economic Review, 99*(1): 112-45.

McCaffrey, D. F., Lockwood, J. R., Kortez, D. M., Louis, T. A., & Hamilton, L. S. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics, 29*, 67-102

McCaffrey, D.F., Sass, T. R., Lockwood, J.R., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy, 4*(4), 572-606.

McLaughlin, M. W. & Talbert, J. E. (2001). *Professional communities and the work of high school teaching*. Chicago, IL: University of Chicago Press.

Nye, B., Konstantopoulos, S., & Hedges, L.V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237–257.

Rivkin, S. G., Hanushek, E. A., & Kain, J.F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417–58.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review, 94*(2), 247–52.

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *The Quarterly Journal of Economics, 116*(2). 681-704.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? *Handbook of the Economics of Education, 3*, 249-277.

Sacerdote, B. (2014). Experimental and quasi-experimental analysis of peer effects: Two steps forward? *Annual Review of Economics*, *6*(1), 253–272.

Sanders, W. L., & Rivers, J.C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Research Progress Report. Knoxville: University of Tennessee Value-Added Research and Assessment Center.

Siskin, L. S. (1994). Realms of knowledge: Academic departments in secondary schools. London: Routledge.Coburn, C. E., & Russell, J. L. (2008). District policy and teachers' social networks. *Educational Evaluation and Policy Analysis, 30* (3), 203-235.

Stoyanov, A., & Zubanov. N. (2012). Productivity spillovers across firms through worker mobility. *American Economic Journal: Applied Economics*, *4*(2), 168-98.

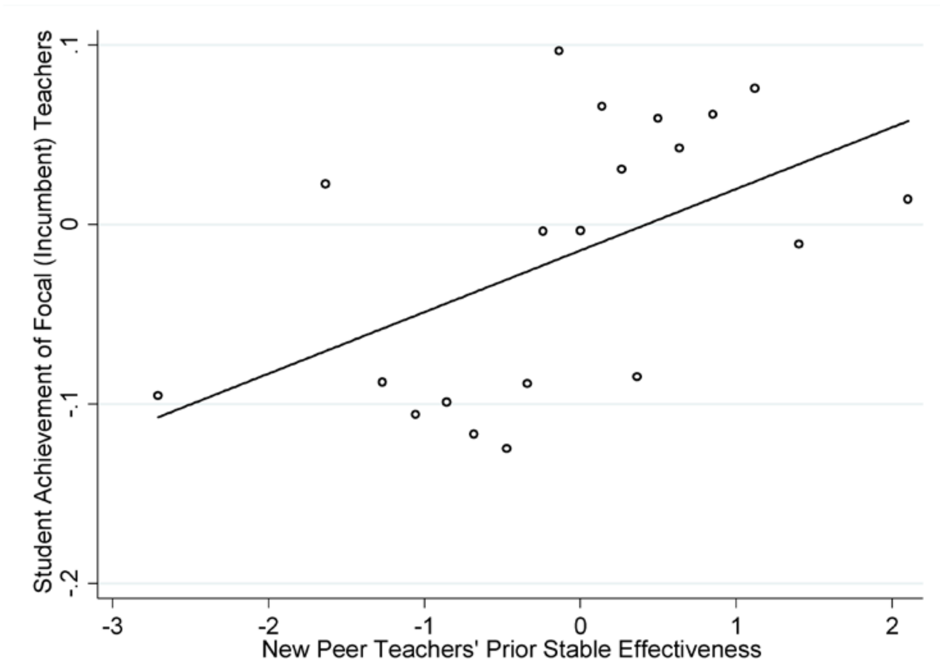
Summers, A.A., & Wolfe, B. L. (1977). Do schools make a difference? *American Economic Review, 6*(4), 639-652.

Xu Z., Özek, U., & Corritore, M. (2012). *Portability of teacher effectiveness across school settings (CALDER Working Paper No.77)*. Washington, D.C.: National Center for Analysis of Longitudinal Data in Education Research, American Institute for Research.

Yuan, K. (2015). A value-added study of teacher spillover effects across four core subjects in middle schools. *Education Policy Analysis Archives, 23*(38), 1-24.

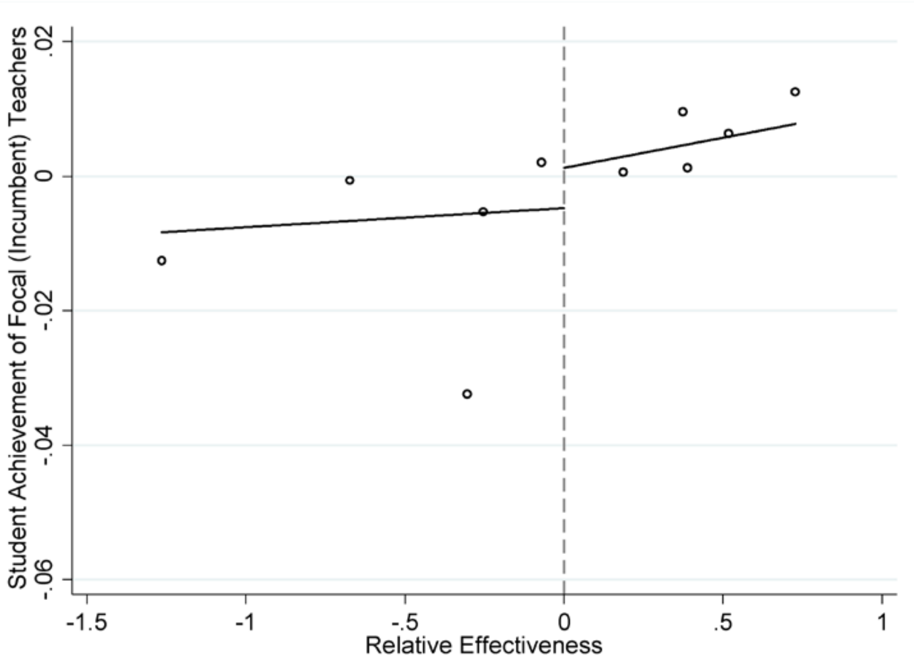
Figures and Tables

Figure 1. "Linear-in-means", the association of new peer teachers' prior stable effectiveness with the student achievement of focal (incumbent) teachers



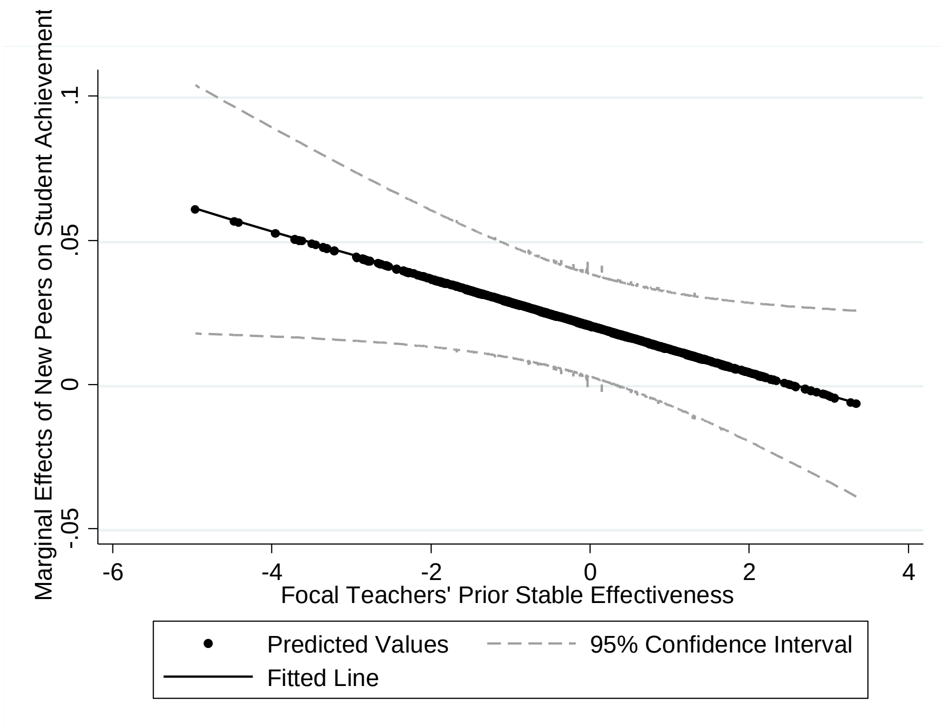
Note. This figure plots the linear relationship between new peer teachers' prior stable effectiveness and the average student achievement of focal (incumbent) teachers, after controlling for focal teachers' own prior stable effectiveness, based on Equation 1.

Figure 2. “Relative effectiveness,” comparing the slope of “relative effective” peers in predicting the student achievement of focal (incumbent) teachers with that of the “relatively ineffective” peers



Note. This figure bases on Equation 2. The x-axis of “relative effectiveness” is the difference in prior stable effectiveness between new peers and the incumbent teacher himself or herself. The linear fit line for cases with the x-axis< 0 is close to flat, which shows a very weak relationship between “relatively ineffective” peers and focal teachers’ student math achievement. In contrast, the linear fit line for cases with the x-axis>0 has a steeper, positive slope, which indicates a much stronger positive association between “relatively effective” peers and focal teachers’ student achievement.

Figure 3. "Absolute effectiveness," the relationship between focal-teacher-specific spillover effects and their own prior stable effectiveness



Note. This figure bases on Equation 3 and plots the marginal effects of new peers on focal teachers' student achievement against focal teachers' prior stable effectiveness. This figure confirms that the spillover is larger for less effective teachers and smaller for more effective teachers. Notably, the spillover effect estimates are negative for a small number of cases and their 95% of the confidence intervals all include zero, indicating that effective focal teachers are not significantly disadvantaged.

Table 1. Descriptive statistics for M-DCPS students and teachers

Variables	Mean	SD
Student-year observations		
Math scores	0.015	0.975
Race/ethnicity: White	0.085	0.280
Black	0.247	0.431
Hispanic	0.647	0.478
Other	0.021	0.142
Female	0.486	0.50
English language learner	0.125	0.330
Eligible for free- and reduced-price lunch (FRL)	0.722	0.448
Special education	0.117	0.321
N	1,150,468	
Transfer teacher-year observations		
Female	0.770	0.421
Race/ethnicity: White	0.368	0.482
Black	0.344	0.475
Hispanic	0.267	0.443
Other	0.021	0.142
Advanced degree (Master's or higher)	0.432	0.495
Standardized value-added	-0.205	0.947
Teaching experience in this school district	7.629	6.614
Total days of absence	5.825	5.613
N	1,594	
Incumbent teacher- year observation		
Female	0.832	0.374
Race/ethnicity: White	0.456	0.498
Black	0.296	0.456
Hispanic	0.237	0.423
Other	0.015	0.121
Advanced degree (Master's or higher)	0.453	0.498
Standardized value-added	-0.046	0.937
Teaching experience in this school district	11.169	8.515
Total days of absence	6.581	6.194
N	26,346	

Table 2. Estimated grade-level spillover effects

	Model-1	Model-2	Model-1	Model-2	Model-1	Model-2
"Linear-in-means": Prior stable effectiveness of new transfer peers	0.010† (0.006)	0.020* (0.009)			0.010 (0.006)	0.021* (0.009)
"Relatively effective": Positive values of the difference in prior stable effectiveness between new transfer peers and focal teachers			0.019*** (0.005)	0.028*** (0.008)		
"Relatively ineffective": Negative values of the difference in prior effectiveness between new transfers and focal teachers			0.002 (0.006)	-0.009 (0.008)		
F-statistics for the difference between "relatively effective" and "relatively ineffective"			3.77†	6.88**		
"Absolute effectiveness": Prior stable effectiveness of new transfer peers × Prior stable effectiveness of focal teachers					-0.006 (0.004)	-0.008* (0.004)
Students' own (focal) teachers' prior effectiveness	0.068*** (0.006)	0.069*** (0.006)	0.081*** (0.008)	0.095*** (0.011)	0.069*** (0.005)	0.069*** (0.005)
N	109,422	109,422	109,422	109,422	109,422	109,422
adj. R-sq	0.687	0.690	0.687	0.690	0.687	0.690

Note: data from 2003-04 to 2012-13.
All models include student and classroom covariates. Student covariates include eligibility for free or reduced price lunch (FRL), whether the student is an English language learner, the student's race/ethnicity, gender, age, prior suspension, prior absence, and prior math and reading test scores. Classroom covariates include % of FRL, % of students are English language learners, % of Hispanic, % of African American, % of Asian, % of White, % of female, average age, average days of prior suspension, average days of prior absence, and the average and standard deviation of students' prior math and reading test scores.
Model-1 includes school-grade, year fixed effects; Model-2 includes school-year, grade fixed effects.
Standard errors are included in the parentheses and clustered at the school-grade-year level.
† p <0.1 * p <0.05 ** p <0.01 *** p <0.001

Table 3. Robustness check on teacher sorting

	With Other New Teachers		Only Same-Grade Teachers	
	Model-1	Model-2	Model-1	Model-2
Panel 1: “Linear-in-means”				
“Linear-in-means”	0.009	0.020*	0.016**	0.032**
	(0.006)	(0.010)	(0.005)	(0.010)
Students’ own (focal) teachers’ prior stable effectiveness	0.069***	0.069***	0.077***	0.082***
	(0.005)	(0.006)	(0.007)	(0.008)
Number of other new teachers (novice teachers or new transfer teachers without prior stable effectiveness) at the grade level	-0.006	-0.002		
	(0.017)	(0.033)		
N	109,422	109,422	56,469	56,469
adj. R-sq	0.687	0.690	0.687	0.690
Panel 2: “Relative effectiveness”				
“Relatively effective”	0.019***	0.028***	0.021**	0.027**
	(0.005)	(0.008)	(0.007)	(0.010)
“Relatively ineffective”	0.002	-0.009	-0.003	-0.023*
	(0.006)	(0.008)	(0.006)	(0.010)
F-statistics	3.658†	6.56**	7.75**	9.9**
Students’ own (focal) teachers’ prior stable effectiveness	0.081***	0.095***	0.092***	0.115***
	(0.008)	(0.011)	(0.009)	(0.014)
Number of other new teachers (novice teachers or new transfer teachers without prior stable effectiveness) at the grade level	-0.005	-0.002		
	(0.017)	(0.033)		
N	109,422	109,422	56,469	56,469
adj. R-sq	0.687	0.690	0.686	0.690
Panel 3: “Absolute effectiveness”				
“Linear-in-means”	0.010	0.021*	0.014*	0.032**
	(0.006)	(0.009)	(0.006)	(0.010)
“Absolute effectiveness”:	-0.006	-0.008*	-0.009*	-0.008†
New transfer peers × Focal teachers	(0.004)	(0.004)	(0.004)	(0.004)
Students’ own (focal) teachers’ prior stable effectiveness	0.069***	0.069***	0.075***	0.081***
	(0.005)	(0.005)	(0.007)	(0.008)
Number of other new teachers (novice teachers or new transfer teachers without prior stable effectiveness) at the grade level	-0.006	-0.003		
	(0.017)	(0.033)		
N	109,422	109,422	56,469	56,469
adj. R-sq	0.687	0.690	0.686	0.690

Note: data from 2003-04 to 2012-13. † p <0.1 * p <0.05 ** p <0.01 *** p <0.001

All models include student and classroom covariates. Student covariates include eligibility for free or reduced price lunch (FRL), whether the student is an English language learner, the student’s race/ethnicity, gender, age, prior suspension, prior absence, and prior math and reading test scores. Classroom covariates include % of FRL, % of students are English language learners, % of Hispanic, % of African American, % of Asian, % of White, % of female, average age, average days of prior suspension, average days of prior absence, and the average and standard deviation of students’ prior math and reading test scores.

Model-1 includes school-grade, year fixed effects; Model-2 includes school-year, grade fixed effects.

Standard errors are included in the parentheses and clustered at the school-grade-year level.

Table 4. Estimated spillover effects of involuntary transfers

	(1)	(2)	(3)
"Linear-in-means": Prior stable effectiveness of new transfer peers	0.015 (0.017)		0.019 (0.018)
"Relatively effective": Positive values of the difference in prior stable effectiveness between new transfer peers and focal teachers		0.031 (0.022)	
"Relatively ineffective": Negative values of the difference in prior effectiveness between new transfers and focal teachers		-0.009 (0.014)	
F-statistics for the difference between "relatively effective" and "relatively ineffective"		1.99	
"Absolute effectiveness": Prior stable effectiveness of new transfer peers × Prior stable effectiveness of focal teachers			-0.014 (0.025)
Students' own (focal) teachers' prior effectiveness	0.038* (0.014)	0.066* (0.025)	0.038* (0.014)
N	5,644	5,644	5,644
adj. R-sq	0.663	0.663	0.663

Note: data from 2019-10, 2010-11, 2011-12
All models include student, classroom, and school covariates. Student covariates include eligibility for free or reduced price lunch (FRL), whether the student is an English language learner, the student's race/ethnicity, gender, age, prior suspension, prior absence, and prior math and reading test scores. Classroom and school covariates include % of FRL, % of students are English language learners, % of Hispanic, % of African American, % of Asian, % of White, % of female, average age, average days of prior suspension, average days of prior absence, and the average and standard deviation of students' prior math and reading test scores.
Grade and year fixed effects are included.
Standard errors are included in the parentheses and clustered at school-grade-year level.
† p <0.1 * p <0.05 ** p <0.01 *** p <0.001

Table 5. Estimated grade-level spillover effects in elementary and secondary schools separately

	Elementary						Middle					
	Model 1	Model 2	Model 1	Model2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
“Linear-in-means”	0.021*	0.001			0.022*	0.000	0.005	0.021+			0.005	0.022+
	(0.008)	(0.018)			(0.009)	(0.018)	(0.007)	(0.013)			(0.007)	(0.012)
“Relatively effective”			0.026**	0.013					0.016*	0.029**		
			(0.008)	(0.014)					(0.007)	(0.010)		
“Relatively ineffective”			-0.008	0.011					0.006	-0.010		
			(0.010)	(0.017)					(0.008)	(0.011)		
F-statistics			6.45*	0.01					0.94	4.50*		
“Absolute effectiveness”: New transfer peers × Focal teachers					-0.010	-0.012†					-0.004	-0.006
					(0.006)	(0.007)					(0.004)	(0.004)
Students’ own (focal) teachers’ prior effectiveness	0.090***	0.091***	0.112***	0.091***	0.089***	0.088***	0.052***	0.054***	0.064***	0.082***	0.055***	0.054***
	(0.008)	(0.008)	(0.011)	(0.020)	(0.008)	(0.008)	(0.007)	(0.007)	(0.010)	(0.015)	(0.007)	(0.007)
N	41,178	41,178	41,178	41,178	41,178	41,178	68,244	68,244	68,244	68,244	68,244	68,244
adj. R-sq	0.666	0.671	0.666	0.671	0.666	0.671	0.698	0.700	0.697	0.700	0.697	0.700

Note: data from 2003-04 to 2012-13.
All models include student and classroom covariates. Student covariates include eligibility for free or reduced price lunch (FRL), whether the student is an English language learner, the student’s race/ethnicity, gender, age, prior suspension, prior absence, and prior math and reading test scores. Classroom covariates include % of FRL, % of students are English language learners, % of Hispanic, % of African American, % of Asian, % of White, % of female, average age, average days of prior suspension, average days of prior absence, and the average and standard deviation of students’ prior math and reading test scores.
Model-1 includes school-grade, year fixed effects; Model-2 includes school-year, grade fixed effects.
Standard errors are included in the parentheses and clustered at the school-grade-year level.
† p <0.1 * p <0.05 ** p <0.01 *** p <0.001

Appendix A

Table A1 provides the estimates for the four types of spillover effects on students' reading achievement. The different model specifications test the robustness of results to teacher selection issues as discussed in the section of "Teacher Sorting". Overall, the "linear-in-means", "relatively effective", and "relatively ineffective" effects are indistinguishable from zero. This absence of spillover effects on reading test scores is anticipated for at least two reasons. First, teachers generally have a stronger influence on math than reading test scores, as identified in prior studies (e.g., Nye, Konstantopoulos, & Hedges, 2004) and also confirmed by our study in that the estimated effects of own teachers on reading test scores are only about one third to one half of the estimated effects of own teachers on math test scores. Second, one of our prior studies reveals a much stronger positive association of teachers' instructional collaboration with their math value-added than reading value-added in M-DCPS (Author, 2015, pp.28-30).

However, the "absolute effectiveness" model in reading shows similar inferences to those identified in math. With one standard deviation increase in own teachers' prior effectiveness, the spillover effect from new transfer peers would decrease about 0.6 percent or 1 percent of one standard deviation of student test scores. In other words, new peer teachers matter less for students whose own teachers were relatively more effective, or equivalently, that they matter more for those students whose own teachers were less effective.

Table A1. Estimated spillover effects on reading test scores

	Main Model		**With Other New Teachers**	
	Model-1	Model-2	Model-1	Model-2
Panel 1: "Linear-in-means"				
"Linear-in-means"	-0.007†	0.003	-0.006	0.003
	(0.004)	(0.007)	(0.004)	(0.007)
Students' own (focal) teachers' prior stable effectiveness	0.028***	0.033***	0.028***	0.033***
	(0.004)	(0.004)	(0.004)	(0.004)
Number of other new teachers (novice teachers or new transfer teachers without prior stable effectiveness) at the grade level			-0.010	-0.011
			(0.011)	(0.013)
Own teacher was in a different grade last year				
N	110,587	110,587	110,587	110,587
adj. R-sq	0.660	0.662	0.660	0.662
Panel 2: "Relative effectiveness"				
"Relatively effective"	-0.002	0.008	-0.002	0.008
	(0.004)	(0.005)	(0.004)	(0.005)
"Relatively ineffective"	0.007†	0.006	0.007†	0.006
	(0.004)	(0.005)	(0.004)	(0.007)
F-statistics	2.35	0.02	2.29	0.03
Students' own (focal) teachers' prior stable effectiveness	0.023***	0.034***	0.023***	0.034***
	(0.005)	(0.008)	(0.005)	(0.008)
Number of other new teachers (novice teachers or new transfer teachers without prior stable effectiveness) at the grade level			-0.010	-0.011
			(0.011)	(0.013)
Own teacher was in a different grade last year				
N	110,587	110,587	110,587	110,587
adj. R-sq	0.660	0.662	0.660	0.662
Panel 3: "Absolute effectiveness"				
"Linear-in-means"	-0.006	0.002	-0.006	0.002
	(0.004)	(0.007)	(0.004)	(0.007)
"Absolute effectiveness":	-0.006†	-0.010**	-0.006†	-0.010**
New transfer peers × Focal teachers	(0.003)	(0.003)	(0.003)	(0.003)
Students' own (focal) teachers' prior stable effectiveness	0.029***	0.033***	0.029***	0.033***
	(0.004)	(0.004)	(0.004)	(0.0043
Number of other new teachers (novice teachers or new transfer teachers without prior stable effectiveness) at the grade level			-0.010	-0.009
			(0.011)	(0.013)
Own teacher was in a different grade last year				
N	110,587	110,587	110,587	110,587
adj. R-sq	0.660	0.662	0.660	0.662

Note: data from 2003-04 to 2012-13. † p <0.1 * p <0.05 ** p <0.01 *** p <0.001
All models include student and classroom covariates. Student covariates include eligibility for free or reduced price lunch (FRL), whether the student is an English language learner, the student's race/ethnicity, gender, age, prior suspension, prior absence, and prior math and reading test scores. Classroom covariates include % of FRL, % of students are English language learners, % of Hispanic, % of African American, % of Asian, % of White, % of female, average age, average days of prior suspension, average days of prior absence, and the average and standard deviation of students' prior math and reading test scores.
Model-1 includes school-grade, year fixed effects; Model-2 includes school-year, grade fixed effects.
Standard errors are included in the parentheses and clustered at the school-grade-year level.

Appendix B

Equation B1 describes the teacher value-added model, which predicts the achievement of student *i*, taught by teacher *j* in school *s* in year *t* as a function of his/her prior achievement, time-invariant and time-varying student characteristics, classroom characteristics, time-varying school characteristics, and a teacher-by-year fixed effect.

$$A_{ijst} = \alpha_0 + \alpha_1 A_{ijst-1} + \alpha_2 A^{other}_{ijst-1} + \boldsymbol{\gamma_1 X}_{ijst} + \boldsymbol{\gamma_2 C}_{jst} + \boldsymbol{\gamma_3 S}_{st} + \theta_{jt} + \varepsilon_{ijst} \quad \text{(B1)}$$

where A_{ijst} is the math or reading exam score of student *i*, taught by teacher *j*, in school *s* and year *t*. The test scores used to generate the value-added estimates are the scaled scores from the Florida Comprehensive Assessment Test (FCAT), standardized to have a mean of 0 and a standard deviation of 1 for each grade in each year. Superscripts of subjects are omitted for simplicity, but we estimate Equation B1 separately in math and reading. $A_{ijgst-1}$ indicates this student's prior year subject test score and $A^{other}_{ijgst-1}$ indicate his or her prior year score in the other subject (e.g., if modeling math achievement, then reading would be the other subject).

$\mathbf{X}_{ijst}$ is a vector of student *i*'s characteristics, including poverty status, whether the student is an English language learner or in special education programs, the student's race, gender, and prior absence. $\mathbf{C}_{jst}$ is a vector of classmates' characteristics, such as percent of students eligible for subsidized lunch, percent of students who are English language learners or in special education programs, percent of Hispanic, African American, Asian, and other students, percent of female students, average prior scores in math and reading, and average days absent. $\mathbf{S}_{st}$ is a vector of school characteristics, including percent of students eligible for subsidized lunch, percent of students who are English language learners or in special education programs, percent of Hispanic, African American, Asian, and other students, percent of female students, average prior scores in math and reading, and average days absent.

θ_{jt} reflects the contribution of a given teacher to student achievement in each year, after controlling for all observed time-varying student, classroom, and school characteristics, and time-invariant student characteristics that may be associated with learning. Since we use the grades (3 to 8) as our reference groups, the estimates also indicate a teacher's deviation from the average teacher in the grade.

After estimating Equation B1, we shrink the teacher-by-year fixed effect estimates using the empirical Bayes methods to adjust for sampling and measurement errors and bring imprecise estimates closer to the mean (see Author, 2012, for a description of the shrinking). After shrinking the value-added estimates, we standardize them to have a mean of 0 and a standard deviation of 1 in each year to facilitate interpretation.

Appendix C

To test the stability of grade-level spillover patterns on students' math achievement, we expand the definition of peers to all teachers who taught in the same school in the same year. We anticipate similar directions of spillover effects, but a decrease in the magnitude from the grade-level effects, because a teacher shares less of the common production processes and direct interactions with peers schoolwide than those taught the same grade. As aforementioned, the identification strategy for schoolwide spillover may suffer from school-year specific shocks that cannot be captured by controlling for student characteristics and their school-average performance. We thus alert the readers to the weaker internal validity.

The "linear-in-means" effects remain positive but significant, as shown in Table C1. The magnitude of "linear-in-means" effects drops to 0.003 from 0.006 of the corresponding grade-level estimates. The "relatively effective" estimate is consistently significantly positive, with an estimate of 1.7 percent of standard deviation increase in student achievement if the prior stable performance of new transfers is one standard deviation higher than that of own teacher. The "relatively ineffective" estimate is 0.004 and nonsignificant, similar as grade-level estimates. The "absolute effectiveness" estimate is -0.003, but insignificant. Overall, the school level models give somewhat similar effects even with some potential bias.

Salient disparities of teacher spillovers are observed between elementary and middle grades. As shown in Table C2, the "linear-in-means" estimate at elementary grades is about 0.015, significant at the 0.05 level, while the middle school estimates are close to zero and not statistically significant. Similarly, the "relatively effective" estimate for elementary grades is 0.024 and statistically significant, while the estimates for middle school are, again, small and not distinguishable from zero. Given the worries with bias in the school level model, we do not put too much weight on these differences.

Table C1. Estimated school-level spillover effects

	(1)	(2)	(3)
“Linear-in-means”: Prior stable effectiveness of new transfer peers	0.003 (0.005)		0.003 (0.005)
“Relatively effective”: Positive values of the difference in prior stable effectiveness between new transfer peers and focal teachers		0.017* (0.007)	
“Relatively ineffective”: Negative values of the difference in prior effectiveness between new transfers and focal teachers		0.004 (0.006)	
F-statistics for the difference between “relatively effective” and “relatively ineffective”		1.95	
“Absolute effectiveness”: Prior stable effectiveness of new transfer peers × Prior stable effectiveness of focal teachers			-0.003 (0.003)
Students’ own (focal) teachers’ prior effectiveness	0.102*** (0.004)	0.101*** (0.004)	0.106*** (0.006)
N	320,715	320,715	320,715
adj. R-sq	0.666	0.666	0.666

Note: data from 2003-04 to 2012-13
All models include student, classroom, and school covariates. Student covariates include eligibility for free or reduced price lunch (FRL), whether the student is an English language learner, the student’s race/ethnicity, gender, age, prior suspension, prior absence, and prior math and reading test scores. Classroom and school covariates include % of FRL, % of students are English language learners, % of Hispanic students, % of African American, % of Asian, % of White, % of female students, average age, average days of suspension, average days of absence, and the average and standard deviation of students’ prior math and reading test scores.
Grade and year fixed effects are included..
Standard errors are included in the parentheses and clustered at the school- year level.
† p <0.1 * p <0.05 ** p <0.01 *** p <0.001

Table C2. Estimated school-level spillover effects in elementary and secondary schools separately

	Elementary			Middle		
"Linear-in-means": Prior stable effectiveness of new transfer peers	0.015* (0.006)		0.015* (0.006)	-0.002 (0.005)		-0.002 (0.005)
"Relatively effective": Positive values of the difference in prior stable effectiveness between new transfer peers and focal teachers		0.024* (0.010)			0.006 (0.008)	
"Relatively ineffective": Negative values of the difference in prior effectiveness between new transfers and focal teachers		-0.010 (0.008)			0.005 (0.006)	
F-statistics for the difference between "relatively effective" and "relatively ineffective"		7.45**			0.00	
"Absolute effectiveness": Prior stable effectiveness of new transfer peers × Prior stable effectiveness of focal teachers			-0.005 (0.004)			-0.003 (0.003)
Students' own (focal) teachers' prior effectiveness	0.121*** (0.006)	0.136*** (0.008)	0.120*** (0.006)	0.079*** (0.005)	0.078*** (0.007)	0.078*** (0.005)
N	129,915	129,915	129,915	190,800	190,800	190,800
adj. R-sq	0.642	0.642	0.642	0.684	0.684	0.684

Note: data from 2003-04 to 2012-13
All models include student, classroom, and school covariates. Student covariates include eligibility for free or reduced price lunch (FRL), whether the student is an English language learner, the student's race/ethnicity, gender, age, prior suspension, prior absence, and prior math and reading test scores. Classroom and school covariates include % of FRL, % of students are English language learners, % of Hispanic students, % of African American, % of Asian, % of White, % of female students, average age, average days of suspension, average days of absence, and the average and standard deviation of students' prior math and reading test scores.
Grade and year fixed effects are included.
Standard errors are included in the parentheses and clustered at the school- year level.
† p<0.1 * p <0.05 ** p <0.01 *** p <0.001