Teacher Turnover, Teacher Quality and Student Achievement in DCPS

Melinda Adnot
Thomas Dee
Veronica Katz
James Wyckoff

Teacher Turnover, Teacher Quality and Student Achievement in DCPS

Melinda Adnot
University of Virginia

Thomas Dee
University of Virginia

Veronica Katz
University of Virginia

James Wyckoff
University of Virginia

Contents

Acknowledgements

Teacher Turnover, Teacher Quality and Student Achievement in DCPS
Melinda Adnot, Thomas Dee, Veronica Katz, James Wyckoff
CALDER Working Paper No. 153
March 2016

Abstract

In practice, teacher turnover appears to have negative effects on school quality as measured by student performance. However, some simulations suggest that turnover can instead have large, positive effects under a policy regime in which low-performing teachers can be accurately identified and replaced with more effective teachers. This study examines this question by evaluating the effects of teacher turnover on student achievement under IMPACT, the unique performance-assessment and incentive system in the District of Columbia Public Schools (DCPS). Employing a quasi-experimental design based on data from the first year years of IMPACT, we find that, on average, DCPS replaced teachers who left with teachers who increased student achievement by 0.08 SD in math. When we isolate the effects of lower-performing teachers who were induced to leave DCPS for poor performance, we find that student achievement improves by larger and statistically significant amounts (i.e., 0.14 SD in reading and 0.21 SD in math). In contrast, the effect of exits by teachers not sanctioned under IMPACT is typically negative but not statistically significant.

I. Introduction

Having an effective teacher can dramatically alter students' educational and economic outcomes. Yet, we know there are substantial differences in the quality of public school teachers, and increasing evidence that in some urban areas less effective teachers are often concentrated in lower-performing schools serving disadvantaged students. Policymakers and researchers recognize these issues and have sought policies to provide all children with effective teachers. The selective retention of effective teachers has been one of the most-discussed strategies that may contribute to this goal. In theory, districts could dismiss ineffective teachers, hire more effective teachers and redouble efforts to retain effective teachers in these schools. However, we know relatively little about how such policies would work in practice. In particular, the capacity of districts to identify effective teachers at the hiring stage is limited (Boyd, et al., 2008; Rockoff & Speroni, 2010; Rockoff, et al., 2011). Furthermore, research and practice have only recently begun making progress on accurately and reliably assessing teacher effectiveness. Some simulations (Hanushek, 2009; Staiger & Rockoff, 2010) estimate that dismissal of the least effective teachers would dramatically improve student achievement. However, these simulations make assumptions regarding the retention of more effective teachers and the labor supply of new teachers that may be overly optimistic. For example, if teacher evaluation and dismissal policies cause more effective teachers to feel their jobs are threatened, such policies may have the unintended consequence of actually lowering teacher quality (Rothstein, 2015). The ultimate outcome depends on the details of the policy, the behavioral response of teachers, and the characteristics of the local labor market from which new teachers are hired.

Some school districts have begun to implement rigorous teacher-evaluation policies that could systematically dismiss meaningful numbers of ineffective teachers (Thomsen, 2014). However, we are unaware of any research that documents how the patterns of teacher turnover created by such policies (i.e., the attrition of teachers sanctioned for low performance, other teachers choosing to leave, and the hiring of new teachers) influence student achievement. In this study, we provide such evidence by examining the effects of teacher turnover under IMPACT, a seminal teacher-evaluation and compensation system introduced in the District of

Columbia Public Schools (DCPS). Implemented at the beginning of the 2009-10 school year, IMPACT evaluates all teachers annually based on multiple measures of effectiveness. Teachers rated as "Highly Effective" are offered large financial and non-financial rewards; those rated as "Ineffective" or twice consecutively "Minimally Effective" are separated from the district. In a purely descriptive sense, DCPS appears to be successful employing compositional change to improve the quality of teaching. The average IMPACT scores of entering teachers exceed those of exiting teachers in each of the three years for which we have data (Figure 1); average differences vary between a third and a half of a standard deviation of teacher quality. Although these are impressive differences, there are a variety of reasons why the overall averages may misrepresent the ability of DCPS to systematically improve teacher quality in the classrooms where they are most needed.

We employ a quasi-experimental event study to examine teacher turnover and its effect on student achievement in DCPS in 2011 through 2013. Specifically, we rely on school-grade-year cells as our fundamental unit of observation and examine, in "difference-in-differences" specifications, how the patterns of teacher mobility influence student test performance in math and English language arts (ELA). We find that teacher turnover in DCPS had an overall positive effect on student achievement in math (i.e., 0.08 standard deviations), and that the effect of turnover in reading is also positive (i.e., 0.046 standard deviations) but is only significant at the 10 percent level. However, the overall effect of teacher turnover masks considerable heterogeneity. We find that, when low-performing teachers (i.e., those with "Ineffective" or "Minimally Effective" ratings under IMPACT) leave the classroom, student achievement grows by 21 percent of a standard deviation in math and 14 percent of a standard deviation in reading. We also find that the attrition of high-performing teachers (i.e., those rated "Effective" or "Highly Effective") has a negative but statistically insignificant effect on student performance.

To be clear, this paper should not be viewed as an evaluation of IMPACT, or even as an assessment of IMPACT's differential effect on teacher composition in DCPS.1 Instead, we believe this paper makes an important contribution by examining the effects of teacher turnover under a unique policy regime. The existing literature

---

[1] A conventional impact evaluation is not feasible because IMPACT was a district-wide reform and because data from the pre-IMPACT era are not generally available. However, the regression-discontinuity study by Dee and Wyckoff (2015) indicates how the incentive contrasts created by IMPACT influenced the retention and subsequent performance of extant teachers.

finds that teacher turnover negatively influences student achievement (Bryk & Schneider, 2002; Guin, 2004;Ronfeldt, Loeb, & Wyckoff, 2013), perhaps due to the difficulty of replacing experienced teachers who leave and through disruptive effects among the teachers who remain. However, teacher turnover may instead have large positive effects if school districts can accurately and systematically identify low-performing teachers and replace them with more effective teachers. But whether that can be achieved at scale in a real-world setting is an open, empirical question. In our study of DCPS schools under IMPACT, we find overall positive effects of teacher turnover. However, these effects are highly heterogenous. The exits of teachers identified as low-performing on average meaningfully improve student achievement, while in some cases exits of high-performing teachers negatively influence achievement. Critically, the supply of entering teachers appears to be of sufficient quality to sustain a relatively high turnover rate. Nonetheless, retaining more high-performing teachers would provide substantial direct and indirect benefits.

II. Background

Improving teacher quality in schools with poor, low-performing and largely non-white students has become an imperative of education policy. A recent body of research has made clear that the variance in teacher effectiveness is qualitatively large and that more effective teachers can dramatically improve students' short and long-run life outcomes (Aaronson, Barrow, & Sander, 2007; Chetty et al., 2014; Jackson, 2012; Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004). Furthermore, the evidence of stark inequities in access to effective teachers (Goldhaber et al., 2015; Isenberg et al., 2013; Sass et al., 2012) has also motivated efforts to improve teacher effectiveness as a means of reducing educational and economic inequality. Policies to improve teacher effectiveness can be conceptualized as either improving the performance of existing teachers or altering the composition of teachers. In this study, we focus on how changes in the composition of the teacher workforce (i.e., due to turnover and hiring) influences student outcomes under a system of performance assessment.[2]

---

[2] In reality, teacher policies often operate through both margins. That is, policies intended to alter the composition of the workforce may also induce existing teachers to improve. Indeed, the design of IMPACT is intended to operate in this fashion.

Composition of the Teaching Workforce

The composition of a district's teachers improves when their policies retain the most effective teachers, exit poorly performing teachers and select the most able entering teachers. High-performing teachers leave their schools and districts for a variety of reasons, some personal, but most related to attributes of their jobs (Johnson, Kraft, & Papay, 2011). Several descriptive studies link teacher turnover to negative school environments and poor student outcomes (Bryk & Schneider, 2002; Guin, 2004). And this teacher turnover is likely to further exacerbate poor school performance through several mechanisms. Quasi-experimental evidence from New York City finds that teacher turnover leads to lower student achievement (i.e., -0.08 SD in math and -0.05 SD in reading), and that some of that effect reflects the disruptive nature of turnover on the students of teachers who remain (Ronfeldt, et al., 2013). This evidence suggests that, in the absence of policies that effectively improve the composition of teachers, we should expect turnover to result in a decrease in student achievement.

Increasing the retention of effective teachers would appear to be an obvious strategy to improve teaching effectiveness, yet over a third of high-performing teachers report that they received little encouragement from their principals to remain at their current school (TNTP, 2012). There is only limited evidence that financial incentives make a difference in retaining teachers generally (Clotfelter, Glennie, Ladd, & Vigdor, 2008; Glazerman & Seifullah, 2012) and high-performing teachers, specifically (Dee & Wyckoff, 2015). Teacher tenure decisions offer an opportunity to differentially retain the most effective novice teachers. While states are increasingly employing more rigorous evaluations as part of tenure reviews, nearly all teachers reviewed are granted tenure (National Council on Teacher Quality, 2012). A tenure review process that more meaningfully differentiates teacher effectiveness is associated with substantial improvements in teacher quality and student achievement (Loeb, Miller, & Wyckoff, 2015).

In the absence of real-world evidence on the effects of policies that improve teacher composition, researchers have simulated the effects of such policies. It is estimated that annually replacing teachers who fall in the bottom 5 to 10 percent of the value-added distribution would improve student achievement by 50 percent of a standard deviation (Hanushek, 2009). Similarly, Staiger and Rockoff (2010) suggest that replacing 80 percent of

first-year teachers with new hires would increase average teacher value-added by 8 percent of a standard deviation of student achievement. However, making alternative assumptions regarding the reliability of teacher quality measures and teachers' behavioral responses to retention policies can lead to different outcomes. Districts whose evaluation system leads to the dismissal of meaningful numbers of teachers may face a limited supply of high-quality teachers. Teachers may find the stress and uncertainty of these working conditions outweigh the benefits, including compensation. As a result, such policies will need to be accompanied by improved working conditions or increased compensation (Rothstein, 2015).

In sum, there is at best limited empirical evidence of the effects of differential retention policies on teacher quality and student achievement. What evidence does directly bear on this issue are simulations dependent on a series of simplifying assumptions about the policies and the behavioral responses of existing teachers and the available labor market. Our study leverages the implementation of IMPACT, the high-stakes teacher evaluation and compensation system in DCPS, to examine this issue directly in an at-scale setting.

Teacher Evaluation in DCPS

In just the last few years, the design and implementation of teacher evaluation has evolved quickly as many districts look to improve teacher performance, partly under the encouragement of federal policies such as waivers from the No Child Left Behind (NCLB) Act and the Teacher Incentive Fund (TIF). While these policy innovations are still a work in progress, best-practice principles of effective evaluation are beginning to emerge (Donaldson & Papay, 2015). The District of Columbia Public Schools (DCPS) were an early and influential adopter. DCPS began evaluating teachers under IMPACT, a new performance-assessment and incentive system, during the 2009-10 school year. The design of IMPACT appears consistent with virtually all of the emerging best-practice principles. First, all teachers are evaluated on a multi-faceted measure of teacher performance (e.g., clearly described standards, the use of multiple teacher observations made by different observers and the use of student outcomes). Second, these evaluations are linked to high-powered incentives that include the potential dismissal of low-performing teachers and very large financial incentives for high-performers. Third, in addition to the

feedback associated with the evaluations, teachers are provided with various supports, including instructional

coaching to assist in improving their teaching practice. Because IMPACT has been implemented at scale over a

sustained period, we have an opportunity to observe behavioral responses to the policy.

Each year, teachers receive a final IMPACT score that determines their IMPACT rating and their associated

consequences. IMPACT scores range from 100 – 400 points and are a weighted average of five components

employed to evaluate teachers. During the period of our study, IMPACT scores were translated into one of four

IMPACT ratings, which dictated consequences as shown in Table 1.[3] Specifically, "Ineffective" teachers are

separated from the district, as are teachers who receive two consecutive "Minimally Effective" ratings. The

financial incentives available to high-performing teachers through IMPACT*plus* include one-time bonuses of up to

$25,000[4] and permanent increases to base-pay of up to $27,000 per year, the present value of which are worth up

to $185,259 in current dollars.[5] These design features of IMPACT create sharp incentive contrasts for teachers

with scores local to the ME/E threshold (i.e., dismissal threats) and the HE/E threshold (i.e., the possibility of a

permanent base-pay increase). An earlier study, employing a regression-discontinuity design, shows that, once

these incentives become politically credible, they meaningfully increased the likelihood that teachers rated as ME

exited DCPS (Dee & Wyckoff, 2015).

---

[3] In 2012-13, DCPS introduced a fifth IMPACT rating category (Developing) and increased the minimum score required for a Minimally Effective (ME) rating. These revisions are inconsequential for our study as they influenced consequences for the following school year.

[4] Bonuses are available to teachers who receive a rating of Highly Effective (HE). The size of the bonuses varied based on whether the teacher taught in a high-poverty school (defined to be a school where the percentage of free and reduced price lunch eligible students was at least 60 percent), whether the teacher was in Group 1 (teachers with value-added scores), and whether the teacher taught a high-need subject.

[5] Increases in base-pay are awarded to teachers with two consecutive HE ratings and vary by whether the teacher taught in a high-poverty school and the position of the teacher on the salary schedule.

IMPACT Score	IMPACT Rating	Consequence
100 – 174	Ineffective (I)	Dismissal
175 – 249	Minimally Effective (ME)	Salary not advanced on salary schedule after 1st ME rating; Dismissal after 2nd consecutive ME rating
250 – 349	Effective (E)	None
350 – 400	Highly Effective (HE)	Bonus; Eligible for permanent base pay increase after 2nd consecutive HE rating

Teacher Retention in DCPS

During the period of our analysis, the average DCPS teacher attrition was 18 percent (Figure 2).[6] A recent study of teacher attrition in 16 urban school districts across 7 states finds year-to-year district attrition averages 13 percent but varies between 8 and 17 percent (Papay et al., 2015). This suggests that the annual attrition in DCPS is comparatively high, which may reflect the intended and unintended effects of IMPACT as well as unique features of the local labor market. Interestingly, the attrition rate among teachers rated as "Effective" or "Highly Effective" (high-performers) was only 13 percent. Attrition of these higher-performing teachers was 10 percent in low-poverty schools and 14 percent in high-poverty schools.[7] Some attrition of high-performing teachers undoubtedly results from the same forces that cause attrition in many districts, (e.g., demanding working conditions or unsupportive leadership). However, in DCPS some high-performing teachers may leave *because* they

---

[6] This is the average rate of teacher attrition from 2009-10 to 2011-12. We examine outcomes from 2010-11 to 2012-13 as a function of teacher turnover at the end of 2009-10 to 2011-12.

[7] Prior to 2012-13, schools were identified as high-poverty if more than 60 percent of students were eligible for FRPL. In 2012-13, the threshold was revised down to 50 percent, causing more schools to be identified as high-poverty. Two schools in our sample are identified as high-poverty schools for the first time in 2012-13. However, this change does not affect our results because we examine turnover at the end of the prior year. As such, these two schools are essentially considered low-poverty schools throughout the period of analysis.

find some features of IMPACT demotivating or stressful. If this response is sufficiently large, IMPACT could reduce teacher quality and student achievement. However, IMPACT also leads to the dismissal of ineffective teachers and induces other low-performing teachers to exit voluntarily (Dee & Wyckoff, 2015). The annual attrition rate of low-performing teachers is 46 percent. This implies that low-performing teachers were more than three times as likely to leave as high-performing teachers (Figure 2) and accounted for 36 percent of all exits from DCPS (Figure 3).

This descriptive summary of retention highlights the challenges confronting DCPS to improve student achievement by improving the composition of its teacher workforce. Losing 13 percent of the best teachers each year places strong demands on teacher recruitment to prevent a reduction in achievement in those classrooms. However, exiting 46 percent of low-performing teachers creates substantial opportunity to improve achievement in the classrooms of low-performing teachers. In the remainder of this paper, we explore how teacher turnover in DCPS under IMPACT affects student achievement. We examine this question in the aggregate and separately for low- and high-performing teachers. We also consider whether the relationship between teacher turnover and student achievement varies across schools and over time.

III. Methodology: Conceptual Model & Empirical Strategy

To examine the effects of teacher turnover on student achievement, we employ a panel-based research design that effectively compares how outcomes in school-grade cells changed following the exit of a teacher to the contemporaneous change in school-grade cells where no turnover occurred. We particularly want to understand whether teacher effectiveness and student achievement are higher or lower as a result of exiting teachers. Changes in overall teacher effectiveness depend upon the magnitude of mean difference in effectiveness between entering and exiting teachers and the proportion of teachers who turnover. Changes in student achievement depend on these differences and on the relationship between measured teacher effectiveness and student achievement. Our empirical model attends to these relationships.

To illustrate how our research design utilizes student- and teacher-level data, we begin with the commonly used specification of student-level achievement shown in Equation (1). The achievement of student *i* in

school *s*, grade *g*, assigned to teacher *j* and class *c* during year *t* (A_{isgjct}) is a function of that student's

observables, including prior achievement, (X_{isgjct}), and the attributes of classroom peers, $\bar{X}_{sgjct}$, the teacher's

value-added (μ_{jsgt}), a school fixed effect (π_s), a year fixed effect (τ_t), and an idiosyncratic error term (ε_{isgjct}) that

captures random noise that may occur at the individual and higher levels of aggregation (e.g., school, grade,

classroom):

$$A_{isgjct} = \beta_0 + X_{isgjct}\beta_1 + \bar{X}_{sgjct}\beta_2 + \mu_{jsgt} + \pi_s + \tau_t + \varepsilon_{isgjct} \qquad (1)$$

To control for our student-level covariates, while facilitating further aggregation of this specification, we replace

the dependent variable in equation (1) with the student-level residuals (A^*_{isgjct}) obtained by regressing A_{isgjct} on

X_{isgjct}. Aggregating the resulting equation to the teacher level, we have:

$$\bar{A}^*_{sgcjt} = \beta_0 + \bar{X}_{sgcjt}\beta_2 + \mu_{sgcjt} + \pi_s + \tau_t + \bar{\varepsilon}^*_{sgcjt} \qquad (2)$$

Consider the case of teacher *j* in a particular school and grade who was hired in year *t* to replace teacher *j'* that

left the school and grade at the end of the prior school year. Equation (3) shows the difference in average

achievement of students taught by the entering teacher in the next year compared to that of the exiting teacher

in the prior year.

$$\Delta\bar{A}^*_{sgjt} \overset{\text{def}}{=} \bar{A}^*_{sgjt} - \bar{A}^*_{sgj't-1} = \left(\bar{X}_{sgjt} - \bar{X}_{sgj't-1}\right)\beta_2 + \mu_{sgjt} - \mu_{sgj't-1} + \pi_s - \pi_s + \tau_t - \tau_{t-1} + \bar{\varepsilon}^*_{sgjt} - \bar{\varepsilon}^*_{sgj't-1}$$

$$= \Delta\bar{X}_{sgt}\beta_2 + \Delta\bar{\mu}_{sgjt} + \Delta\tau_t + \Delta\bar{\varepsilon}^*_{sgjt} \qquad (3)$$

That is, equation (3) models the change in student achievement (i.e., conditional on student traits) as a

teacher-level function of the change in classroom peers, the change in teacher quality, and other unobserved

time-varying changes. However, conducting the analysis at the teacher level would have several prohibitive

limitations. For example, if the attrition of a teacher has negative consequences for the productivity of her grade-

level colleagues (Ronfeldt et al. 2013), this specification would not capture it. Furthermore, a teacher-level

specification may exacerbate the bias due to internal-validity threats (Chetty, Friedman, and Rockoff, 2014). For

example, in the presence of teacher turnover, some more motivated parents may seek to have their children

placed in the classrooms of returning teachers, leaving entering teachers with lower-performing students (i.e., net

of observed traits). Such non-random sorting of students to teachers is much less likely at the school-by-grade level, because the sorting would need to occur across schools to affect school-grade outcomes. Furthermore, observations at the annual school-grade level capture spillover effects that may exist among members of a school-grade team. Aggregating to the grade level also avoids the need to match each exiting teacher with the teacher replacing her. The student-weighted aggregation of Equation (3) to the school-grade-year level is shown here:

$$\Delta \bar{A}^*_{sgt} = \Delta \bar{X}_{sgt} \beta_2 + \Delta \bar{\mu}_{sgt} + \omega_t + \Delta \bar{\varepsilon}_{sgt} \qquad (4)$$

Our analysis aims to understand how student achievement changes as a function of teacher turnover, rather than as a function of changes in teacher value-added in equation (4). That is, teacher turnover may change teacher quality (e.g., $\Delta \mu_{sgt}$), which in turn changes student achievement. We estimate a reduced-form model of changes in residualized student achievement as a function of teacher turnover at the end of the prior school year. E_{sgt-1} in Equation (5) is the student-weighted share of teachers in school s and grade g in year t-1 who exit DCPS by the beginning of year t.

$$\Delta \bar{A}^*_{sgt} = \Delta \bar{X}_{sgt} \beta_2 + \gamma_1 E_{sgt-1} + \omega_t + \Delta \varepsilon^*_{sgt} \qquad (5)$$

The identification strategy implied by this research design has a straightforward "difference-in-differences" logic.[8] That is, we effectively examine the *change* in student performance in a school-grade cell before and after teacher turnover has occurred. These student-performance changes reflect both the effect of teacher turnover and the effect of other time-varying determinants. A second difference—the contemporaneous performance change in school-grade cells that did *not* experience turnover—captures the effects of those other time-varying determinants. The difference in these differences then isolates the effect of teacher turnover. Ideally, we would complement this analysis with similar results for the period that preceded the introduction of IMPACT. However, reliable data for teachers linked to students does not exist prior to 2009-10.

Our quasi-experimental specification unrestrictively controls for several unobserved determinants of student achievement. More specifically, this specification identifies the effect of teacher turnover controlling for

---

[8] Also, all estimates cluster standard errors at the school-grade level to account for repeated observations over time.

time-invariant traits specific to each school-grade cell, time-varying traits shared across all schools and grades, and various student-level traits including prior achievement. However, the internal validity of the inferences based on this basic model still rests on several critical assumptions that we engage directly. First, our design implicitly assumes that students don't sort to (or from) turnover classes by switching schools in a way that biases the results. Second, as currently specified, our approach implicitly assumes that, when filling vacancies due to turnovers, schools don't manipulate teacher *transfers* within DCPS in a manner that biases turnover results. For example, when an exit occurs within a school, principals don't systematically move the most or least effective teachers from other grades to fill that vacancy. Although there are slight variations across years and subjects, on average 55 percent of replacement teachers come from outside the DCPS system, 34 percent transfer within DCPS schools and 11 percent transfer across DCPS schools. Our specification also assumes that these teacher transfers have no achievement implications for the "sending" school-grade cell (e.g., due to disruption of teacher peers). Third, our design assumes that there are not important unobserved factors changing at the school or grade-level that influence student achievement and that are also correlated with turnover (e.g., increasingly effective principals).

To address the robustness of our results in the presence of these potential confounds, we modify our basic estimation approach and conduct several robustness tests. First, we add several additional controls to our empirical models to address potential challenges to internal validity. To address concerns that within-school or across-school transfers may influence our results, we also introduce direct controls for these transfers. To assess the relevance of unobserved school trends that are correlated with turnover, we also employ specifications that include school fixed effects. Time-invariant school effects have already been eliminated from our design as a result of first-differencing school-grade observations. Adding a school fixed effect to our first-difference specification implies that we are also controlling flexibly for school-specific *changes* over time (e.g., school trends in culture and leadership).[9] Second, we also estimate auxiliary regressions to examine whether our treatment

---

[9] Controlling for school fixed effects in our first-differenced specification has some disadvantages. These include a loss of statistical precision and a reliance on the variation within schools that have more turnover during our brief sample period.

explains changes in observed student attributes. To the extent that teacher turnover predicts changes in students traits, it would suggest endogenous switching based on unobserved correlates of these observed traits.

Third, we estimate the effects of teacher turnover on teacher quality directly. If, as our conceptual model suggests, teacher quality is the mechanism through which turnover influences student achievement, we should observe consistent results for the effects of turnover on both teacher quality and student achievement. To provide increased assurance that any student achievement changes associated with teacher turnover reflect its effects on teacher quality, we estimate some specifications where we employ IMPACT scores as the dependent variable.[10],[11] As will be seen in Tables 3-5, in every instance where turnover is estimated to positively or negatively affect achievement, we observe an effect of turnover on observed teacher quality that is of the same sign and usually of proportionate magnitude. This increases our confidence that the change in teacher quality is the primary factor determining the effect of turnover on student achievement, and not other factors that could plausibly be associated with teacher exits.

We create three treatment variables to examine different types of teacher turnover. As before E_{sgt-1} is the proportion of students in a school-grade-year cell in the prior year whose teacher left the district and is used in specifications in which we examine the overall effects of turnover. In other specifications, E^L_{sgt-1} denotes the proportion of students in each such cell whose teacher exited DCPS and was a low-performing (i.e., Minimally Effective or Ineffective) teacher, and E^H_{sgt-1} denotes the proportion of students taught by a high-performing (Effective or Highly Effective) teacher who left the district at the end of year *t-1*. In all specifications, we condition on the prevalence of within-school transfers, S_{sgt-1}, and transfers across schools in the district, D_{sgt-1}. These controls allow us to condition on the effects that turnover may have on school-grade cells that "send" teachers elsewhere within the district. The resulting specification takes the following form:

---

[10] Dividing γ_l from Equation 6a by δ_l from Equation 6b approximates the Wald estimator, which represents the change in student achievement due to changes in teacher quality that result from teacher turnover.

[11] There is a debate whether to control for student attributes when examining measures of teacher quality (see, for example, Whitehurst et al., 2014). For teachers in tested grades IMPACT already controls for student characteristics when estimating value added. For other components of IMPACT the logic of student controls is much less obvious. As a result, we omit controls for observable student attributes in equation 6b. This contrasts with our approach to the estimates of student achievement, where there is strong evidence and a long history of controlling for student attributes.

$$\Delta \bar{A}^*_{sgt} = \gamma_1 E^L_{sgt-1} + \gamma_2 E^H_{sgt-1} + \delta S_{sgt-1} + \theta D_{sgt-1} + \Delta \bar{X}_{sgt}\beta_2 + \omega_t + \varphi_s + \varepsilon^*_{sgt} \qquad (6a)$$

$$\Delta \overline{TQ}_{sgt} = \gamma'_1 E^L_{sgt-1} + \gamma'_2 E^H_{sgt-1} + \delta' S_{sgt-1} + \theta' D_{sgt-1} + \omega'_t + \varphi'_s + \varepsilon'_{sgt} \qquad (6b)$$

Finally, we examine whether the effect of teacher turnover varies by year or school characteristics by interacting each treatment variable with the appropriate year or school-characteristic indicator variable (not shown). For instance, we assess whether the effect of teacher turnover differs between high- and low-poverty schools.

IV. DCPS Administrative Data and Sample Construction

Our analysis draws on several sources of student, teacher, and school administrative data from DCPS. Students' test scores, demographic variables, and teacher assignments come from DC's Comprehensive Assessment System (DC CAS). These data span the 2009-10 through 2012-13 school years and include just over 56,564 student-years for tested students in grades 4 through 8 with a prior test scores. Teacher administrative data include annual school assignments and annual IMPACT evaluation data. These data also span the 2009-10 through 2012-13 school years and include almost 1,900 teacher-years for teachers of students with DC CAS achievement data. Finally, school administrative data identify school type, poverty status, and closure status.

To construct our final analytical sample, we edited the data on the students and teachers in several, conventional ways. First, we restricted our sample to general education classrooms, which resulted in dropping 12 special education campuses leaving 103 schools serving students tested in grades 4 through 8. We then excluded students when they were tested in a grade other than their assigned grade (0.22 percent of student-year observations) or when they lacked a prior-year score (1.97 percent of student-year observations). To limit measurement error, we linked teachers to school-grade-year cells if the teacher is assigned to at least ten tested students in that cell. This restriction eliminated 0.62 percent of teacher-school-grade-year observations. We also excluded teacher-year observations when those teachers taught in a school that closed at the end of that school year. This restriction also eliminated 0.62 percent of teacher-school-grade-year observations.

The primary outcome of interest is the year-to-year change in average residualized and standardized student achievement at the school-grade-year level.[12] To construct this measure, we first standardize students' scale scores to have mean zero and unit standard deviation within a subject, grade, and year. Next, we recover the residuals from a student-level linear regression of standardized test scores on lagged standardized test scores and student demographics. Then using the average residuals in each school-grade-year cell, we subtract prior year outcomes from current year outcomes. This measure isolates how test performance in each school-grade cell changed across years after controlling for the prior achievement and outcome-relevant baseline traits of the students it served. We similarly constructed first-differenced measures for the IMPACT scores of teachers in each school-grade-year cell. Because we have achievement data and IMPACT scores for four years (2009-10 to 2012-13), we are able to create three years of these differenced outcomes. Aggregating these observations to school-grade-year cells produces the unrestricted sample, whose descriptive statistics are shown in Table 2 columns 1 (math) and 5 (reading).

A final set of sample restrictions reflects concerns regarding missing data. First, differenced outcomes can only be created when the school-grade cell contains the outcome of interest in two consecutive years. This results in missing observations when schools open or close during the years of our analysis. This restriction produces school-grade-year cells with missing outcome data, which results in a loss of 87 school grade cells in math (838 observations in the unrestricted sample to 751 in the base sample) and 85 in reading (838 to 753). Second, some school-grade-year cells are missing IMPACT scores, which results in different estimation samples for changes in IMPACT scores (equation 6b) versus changes in student achievement (equation 6a). Because we want to observe the effect of teacher turnover on teacher quality and student achievement in the same school-grade-year cells, we drop cells that are missing differenced IMPACT scores. This results in the loss of 17 school-grade-year cells in the math sample and 20 school-grade-year cells in the reading sample. The remaining sample is unbalanced in that each school-grade cell is not observed in each year.

---

[12] See the Technical Appendix for an extended discussion of the sample and variable construction. We create separate math and reading samples because teachers are linked to different students and school-grade-year cells in each subject.

Third, we eliminate school-grade cells with fewer than three years of differenced outcomes. We are concerned that unbalanced observations introduce structural changes that influence estimates in ways that don't reflect responses to typical teacher exits. For example, school-grade-cells may exist in some years but not others because schools close during the time frame of our analysis. In such situations, within-school, time-varying factors which we don't observe, may influence student achievement and be correlated with teacher turnover, biasing our estimates. This restriction results in the loss of an additional 71 school-grade-year cells from the math sample and 67 school-grade-year cells from the reading sample, and creates the balanced sample.

Table 2 summarizes average student and teacher characteristics observed in each of the analytic samples. As might be expected these sample restrictions influence the nature of our sample. Appendix Table 1 statistically compares the differences in the means of the school-grade-year cells that were deleted in moving from the unrestricted to the balanced samples. As might be expected, relative to cells retained in the balanced sample, the dropped cells have lower residualized achievement and a somewhat higher percentage of black students, a lower percentage of Hispanic students and a higher percentage of students attending high-poverty schools. Cells dropped to create the balanced math sample also have lower average IMPACT scores and a higher incidence of teacher turnover. We estimate our basic student achievement model (equation 6a) with and without these observations to explore how these exclusions affect our estimates. As is evident by comparing the estimates in Appendix Tables 2 (math) and 3 (reading), these estimates are similar. In general, these estimates show that, as we restrict the sample, the estimates of the effect of turnover of low-performing teachers become somewhat more positive and the effect of exits of high-performing teachers becomes slightly less negative. Results of models that distinguish between high and low-poverty schools and by year show very similar patterns. These results are available from the authors.

The "treatment" variable in our setting is defined by the proportion of students in a school-grade-year cell experiencing different types of teacher turnover.[13] Teacher-school assignment rosters and rosters that link

---

[13] Teachers on leave of absence were not considered exits in our analysis and thus are not considered in construction of the treatment variable. Also, students in school-grade cells for whom there is insufficient information to include in our analytic

teachers to students allow us to identify teacher exits as well as within and across school transfers. IMPACT ratings distinguish turnover among high-performing teachers (those rated "Effective" or "Highly Effective") from turnover among low-performing teachers (those rated "Ineffective" or "Minimally Effective"). Turnover dosages are calculated by employing teacher-student assignment rosters to identify the proportion of students in school-grade-year cells affected by each type of teacher turnover.

V. Results

Our conceptual model suggests the induced turnover of low-performing teachers (i.e., teachers rated by IMPACT as "Ineffective" or "Minimally Effective") should result in improvements in teaching quality and student achievement, while the turnover of high-performing ("Highly Effective" and "Effective") teachers may well result in a reduction in teacher quality and student achievement depending on the quality of entering teachers. The overall effect, which balances these two types of turnover, is conceptually ambiguous and depends on the composition of exiting teachers and the quality of entering teachers.

Descriptive Summary

Before turning to our estimates, it may be instructive to examine simple averages of the IMPACT scores of exiting and entering teachers. If our estimates, which control for a variety of potential confounds, are wildly different from these simple means, we would want to understand how our adjustments influence the outcomes. Figure 4 shows the unconditional means of IMPACT scores of all exiting and entering general education teachers (i.e., teachers of all subjects in tested *and untested* grades) in DCPS.[14] As might be expected, mean IMPACT scores of exiting high-performing teachers exceed those of entering teachers by 12 to 23 IMPACT points (i.e., 25 to 45 percent of a standard deviation of teacher effectiveness (IMPACT scores)) depending on the year. In contrast,

---

sample were included in the calculation of the treatment variable. We do so to minimize potential bias associated with selective sample attrition.

[14] Our data do not allow us to identify which teachers may fill the specific vacancy left by an exiting teacher. Thus, while we know the IMPACT rating of an exiting teacher, e.g., Ineffective or Minimally Effective, we don't know the IMPACT rating of the teacher who replaced that teacher. Thus, IMPACT scores for entering teachers reflect all entering teachers and not necessary those who replaced an exiting high-performing or low-performing teachers. In the "High-Performing" and the "Low-Performing" panels we employ the overall average for entering teachers.

exiting low-performing teachers are substantially less effective than the average entering teacher, with differences between 71 and 85 IMPACT points (i.e., 1.4 and 1.7 standard deviations). Across all teachers, entering teachers have IMPACT scores between and third and a half of a standard deviation greater than exiting teachers. A very similar pattern exists when the sample is restricted to teachers who can be matched to students with math achievement scores (Figure 5). Here the difference between entering and exiting teachers varies by 35 to 55 percent of a standard deviation of teacher effectiveness depending on the year. The pattern for teachers matched to students with reading scores is identical with somewhat smaller differences between the IMPACT scores of entering and exiting teachers (i.e., 0.25 to 0.50 standard deviations; results available from authors). Importantly for the purposes of our analysis, the differences in Individual Value-Added (IVA) scores for entering and exiting teachers are very similar to those observed for teacher quality (Figure 6). The patterns for the exits of high-performing and low-performing teachers are identical to those observed for IMPACT scores.

Comparing the IMPACT scores of entering and exiting teachers suggests that teacher quality is improving as a result of teacher turnover. This is true whether teacher effectiveness is measured by overall IMPACT scores or by value-added. However, when teachers who are judged to be high-performing voluntarily exit, they are replaced on average by somewhat less effective teachers. Contrast that with the exit of teachers who are either forced to leave as a result of IMPACT or whose performance, if not improved, would lead to a forced exit. Turnover in this instance appears to result in a substantial improvement in measured effectiveness. As discussed above, there are a variety of reasons why these simple comparisons may misrepresent the effects of teacher turnover in DCPS. For example, the composition of students may have changed from one year to the next in a way that either favors or disadvantages teachers entering a school-grade cell which experienced teacher turnover. We now turn to the estimation of equations 6a and 6b, which control for a number of potentially confounding factors.

Quasi-Experimental Estimates

We report our main results (i.e., estimates based on equations 6a and 6b) in Table 3. The results in the first row identify the estimated effect of overall teacher turnover. Interestingly, these results suggest that the exit

of DCPS teachers led to improved teacher quality and student achievement in both math and reading, although the reading estimate is not significant at traditional levels. More specifically, these results imply that, if all students in a school-grade cell experienced turnover of the average exiting teacher IMPACT scores would increase by 17.4 points (Table 3, row 1, column 1). This is approximately a third of a standard deviation (SD) of teacher effectiveness (SD=50). The corresponding increase in student achievement is 0.079 SD. In reading, IMPACT scores are estimated to increase by 15.1 IMPACT points and student achievement is estimated to increase by 0.046 SD, but is only significant at the 0.10 level. Thus, on average exiting teachers are replaced by teachers who are more effective as measured by IMPACT and who increase student achievement, at least in math.

In remaining rows of Table 3, we report the estimates when the effects of teacher turnover are allowed to differ across teacher effectiveness groups (i.e., high and low performers). These results indicate that the overall effects of teacher turnover masked considerable heterogeneity across low- and high-performing teachers. Turnover of high-performing teachers results in a decrease in average IMPACT scores of 30 points (i.e., 0.60 SD of teacher effectiveness) in math. This negative effect reflects the difficulty of replacing a high-performing teacher. Our estimates indicate that turnover of a high-performing teacher has a negative but statistically insignificant effect on student achievement (-0.055 SD). Similar, but smaller, results hold for reading.

In contrast, the exit of low-performing teachers substantially increases both teaching quality and student achievement. In math, the exit of low-performing teachers is estimated to improve teaching quality by 64 IMPACT points (1.3 SD) and student achievement by 0.21 SD. The effects in reading are somewhat smaller but still large, 46 IMPACT points and 0.14 SD of student achievement. Over the first three years of IMPACT, replacing teachers identified by IMPACT as low-performers leads to substantial improvement in student achievement as, on average, their replacements are meaningfully more effective teachers.

Robustness of Results

The consistency of the effects of turnover on teacher quality and student achievement and their robustness to introducing student controls increases our confidence in the internal validity of our estimates.

Nonetheless, legitimate concerns may remain that parents or principals may systematically respond to teacher turnover by altering the assignment of students to teachers in ways that threaten internal validity. For example, if turnover predicts changes in student attributes, it may signal strategic behavior by parents or principals that may bias our results. Fortunately we find nothing of concern when we regress a variety of student characteristics on teacher turnover (Appendix Table 4). Of the 18 estimated coefficients (six student attributes by three types of teachers (all, high-performing and low-performing), only one is significant at conventional levels. The exit of all high-performing teachers from a school-grade cell is associated with a 2.4 percent *decrease* in LEP students. These results suggest that there is not systematic sorting of students to teachers in response to turnover, and when there is some evidence, the magnitudes are modest. Nonetheless, we include controls for all student variables we can observe.

Another potential threat to the validity of our estimates may be that underlying trends in schools may cause student achievement to increase over time in school-grade cells with turnover but not in school-grade cells without turnover. To address this issue, we estimate first-difference models that introduce school fixed effects and models include school-by-year fixed effects. The identifying variation for estimates with school fixed effects comes from within school comparisons of school-grade cells with and without turnover. Adding a school-by-year fixed effect effectively limits our comparisons to grades in the same school and year with and without turnover. Estimates for our base models and those with school and school-by-year fixed effects are shown in Appendix Tables 5 (math) and 6 (reading). Adding school fixed effects to our base model change the estimates only slightly. The one substantive change is the effect of a typical teacher exit on math student achievement. The coefficient is somewhat smaller (0.058 SD rather than 0.79 SD), the standard error larger (0.038 rather than 0.030), the combination of which results in a statistically insignificant estimate. Adding school-by-year fixed effects has a larger effect on some of the estimates. In math, while still significant and educationally meaningful, the effect of turnover on the achievement of low-performers is about half as large as in either of the other two models. In reading the change is not nearly so dramatic. Adding school-by-year fixed effects substantially reduces the identifying variation in ways that have important implications for the identification of effects and for external

19

validity. For example, 663 school-grade-cells contribute to identifying the effects of our preferred specification in math (Table 3). This is reduced to 534 school-grade cells when we include school fixed effects and to only 317 if we include school-by-year fixed effects.

We include one additional robustness check in which we estimate the effects of a "placebo" model where turnover at the end of 2012-13 is used to predict changes in student achievement from 2009-10 to 2010-11. If turnover is the mechanism that drives our results and not some other attribute of the school-grade cells that experience turnover, then the effects of the placebo estimates should not be similar to the estimates presented in Table 3. They are not. As shown in Appendix Table 7 none of the estimated coefficients in math or reading are statistically significant.

Treatment Heterogeneity

There are several other ways in which the effects of teacher turnover may be heterogeneous. For example, the contexts across low and high-poverty schools are likely to shape both the prevalence of teacher turnover and its effects on students. Overall, we find that high-poverty schools appear to improve as a result of teacher turnover. We estimate that the overall effect of turnover on student achievement in high-poverty schools is 0.084 and 0.052 in reading. Both estimates are statistically distinguishable from zero (Table 4, row 2). In comparison, the point estimates of the effect of turnover in low-poverty schools are close to zero.

DCPS appears to be quite capable of replacing exiting high-performing teachers in low-poverty schools with comparable teachers (Table 4, row 3). However replacing a high-performing teacher in a high-poverty school is more difficult and is estimated to result in a decrease of 80 percent of a standard deviation of teacher quality in math and 40 percent of a standard deviation in reading, though corresponding decreases in student achievement are not significant (Tables 4, row 4).

Forty percent of teacher turnover in high-poverty schools is among low-performing teachers (Figure 3). Our estimates indicate that there are consistently large gains from the exit of low-performing teachers in high-poverty schools. In math, teacher quality improves by 1.3 standard deviations and student achievement by 20 percent of a

standard deviation; in reading these figures are 1 standard deviation of teacher quality and 14 percent of standard deviation of student achievement. In DCPS, virtually all low-performing teacher turnover is concentrated in high-poverty schools: on average, 1 percent of students in low-poverty schools experience low-performing teacher turnover.[15]

When we examine the effects of DCPS turnovers over time, we observe substantial consistency as well as a few interesting differences. Overall, the effects of DCPS turnover appear to become increasingly positive year to year. However, student achievement is estimated to be unaffected until 2013 when for math (Table 5, columns 1 and 2, first three rows) and reading (Table 5, columns 5 and 6, first three rows) the estimated effect is an improvement of 11 percent SD of student achievement.

For most years, the exit of high-performing teachers doesn't influence teacher quality or student achievement. However, in one year for math (2012) and reading (2011), the exit of high-performing teachers has a substantial negative effect on teaching quality and student achievement. These estimates are similar across alternative analytic samples that employ the base and unbalanced data. When we examine the individual exiting and entering teachers in the school-grade cells with teacher turnover, we observe the exit of several very effective teachers who are replaced by teachers whose subsequent performance places them among the low-performers.

In contrast, the exit of low-performing teachers yields consistently large improvements in teaching quality and student achievement in math (0.18 to 0.24 SD of student achievement) and increasing effects over time in reading (0.05 (not significant) to 0.21 SD of student achievement). In almost every year DCPS has been able to replace low-performing teachers with high-performing teachers who have been able to improve student achievement.

---

[15] More specifically, 3 low-poverty school-grade-year cells in the math sample experience low-performing teacher turnover and only 1 low-poverty school-grade-year cell in the reading sample experiences low-performing teacher turnover. As a result, we do not present estimates for turnover of low-performing teachers in low-poverty schools.

VI. Discussion

In general, higher rates of teacher turnover are legitimately thought to negatively influence student outcomes (e.g., Ronfeldt et al. 2013). However, DCPS constitutes a unique and policy-relevant case because, under IMPACT, a substantial fraction of teacher turnover consists of lower-performing teachers who were purposefully compelled or encouraged to leave, thus potentially altering the typical distribution of teacher effectiveness among exiting teachers. We find that the overall effect of teacher turnover in DCPS conservatively had no effect on achievement and, under reasonable assumptions, improved achievement. This average combines the negative, but statistically insignificant, effects of exits of high-performing teachers with the very large improvements in student achievement resulting from the exits of low-performing teachers.

The high stakes associated with IMPACT have been controversial, both within the District of Columbia, as well as in broader discussions of education policy. There are elements of both sides of this debate in our estimates. While we are unable to identify high-performing teachers who leave DCPS *because* of IMPACT, our estimates indicate that replacing high-performing teachers who exit with teachers who perform similarly is difficult. In general, such turnover doesn't lead to statistically significant reductions in student performance, except in one notable instance (i.e., math teachers in 2011-12).

Alternatively, IMPACT targets the exit of low-performing teachers. Our estimates show that doing so substantially improves teaching quality and student achievement in high-poverty schools. An improvement of 20 percent of a standard deviation of student achievement in math is roughly equivalent to 35 to 65 percent of a year of student learning, depending on grade level (Hill, Bloom, Black, & Lipsey, 2008). Similarly, improvements of 14 percent of a standard deviation in reading translate to 35 to 55 percent of a year of learning. More than 90 percent of the turnover of low-performing teachers occurs in high-poverty schools, where the proportion of exiting teachers who are low-performers is twice as high as in low-poverty schools. An important component of IMPACT's design is to dismiss teachers rated as "Ineffective" and twice consecutively "Minimally Effective." As is clear from this analysis, the benefits of that policy primarily redound to high-poverty schools. By comparison to

almost any other intervention, these are very large improvements that are situated among some of the neediest students.

We should note that our analysis does not have the causal warrant of an experimental design. Nonetheless, under certain identifying assumptions that we articulate and examine, our quasi-experimental design does identify the change in student achievement caused by teacher turnover. However, we don't claim that IMPACT caused all of the teacher turnover we observe. Although IMPACT certainly caused some teachers to leave DCPS through dismissals, voluntary teacher attrition is likely driven by myriad teacher preferences. While it's possible that teachers may leave DCPS because they are dissatisfied with IMPACT and the human capital strategies in DCPS writ large, we are unable to link the attrition of high-performing teachers to IMPACT.[16] Nor do we know whether our turnover results for teachers and students in grades 4th through 8th in math and reading generalize to turnover for other teachers and students. However, the descriptive summaries in Figures 4 and 5 would suggest they might.

Our empirical results were not inevitable, even for the turnover of low-performing teachers. As Rothstein (2015) makes clear, there are good reasons to believe that the supply of teachers may be insufficient to maintain teacher quality, especially when teacher quality is difficult to ascertain in advance and challenging to improve in schools where there is substantial turnover. Our estimates suggest that, on average, DCPS is able to recruit replacements for exiting teachers who are at least as effective, and for low-performing teachers, replacements who are substantially more effective. These results are consistent with simple descriptive evidence on the effectiveness of entering and exiting teachers (Figures 1, 4 and 5). This may reflect the compensating differentials available to DCPS teachers in the form of bonuses and increases in base pay or it may reflect specific aspects of the market for teachers in the District of Columbia. Other school districts may experience different results when implementing a system intended to increase the attrition of low-performing teachers.

---

[16] As noted earlier, the attrition of high-performing DCPS teachers (i.e., 13 percent) is similar to that observed in other urban districts, suggesting that the per se effect of IMPACT on the attrition of such teachers may not be large.

The challenge of improving the composition of teachers in DCPS is increasing. First, as the least effective teachers exit, there are fewer such teachers to exit over time and we would expect the average effectiveness of exiting teachers to continue to increase. Second, in 2012-13 DCPS adjusted its evaluation system so that to be rated as "Effective" or better (and thus avoid sanctions) teachers needed IMPACT scores of at least 300 rather than 250 as had been true in 2011-12. Increasing the threshold for high-performing status will likely lead to the exit of some previously "Effective" teachers who are now classified as "Developing" and may cause some "Effective" and "Highly Effective" teachers to leave as they perceive the system as more stressful. On the other hand, DCPS made several other changes to IMPACT in 2012-13 which may cause the system to be more hospitable, such as reducing the number of teacher observations, increasing access to bonus and base pay increases and reducing the weight of value-added for Group 1 teachers.

We expect that both the declining numbers of very low-performing teachers and changes in the IMPACT rating thresholds place strong demands on the system to continue recruiting effective teachers to replace the exit of higher-performing teachers. Figure 1 presents some early evidence of these trends. The teachers exiting at the end of our study window were noticeably more effective than those exiting after IMPACT's first year (i.e., by about 40 percent of a teacher-level standard deviation). However, over this same period, the performance of entering teachers also grew appreciably (i.e., 28 percent of a standard deviation). These trends appear unrelated to the average experience of entering and exiting teachers, which, throughout this period, remained relatively constant at 3.5 and 7 years, respectively. As long as DCPS continues to recruit more able teachers than it loses, compositional change will likely lead to increased student achievement. Whether DCPS can reap further performance benefits from compositional change in its workforce as it increases performance standards appears plausible, but remains to be seen. Regardless, our results indicate that, under a robust system of performance assessment, the turnover of teachers can generate meaningful gains in student outcomes, particularly for the most disadvantaged students.

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. http://doi.org/10.1086/508733

Adnot, M., & Wyckoff, J. (2015). Increasing the Effectiveness of Teachers in Low-Performing Schools. In H. Ladd and M. Geortz, (Eds). *Handbook of Research in Education Finance and Policy* (pp. 528-544). New York: Routledge, 2nd Edition.

Atteberry, A., Loeb, S., & Wyckoff, J. (2013). *Do First Impressions Matter? Improvement in Early Career Teacher Effectiveness* (Working Paper No. 19096). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w19096

Barnes, G., Crowe, E., & Schaefer, B. (2007). *The Cost of Teacher Turnover in Five School Districts: A Pilot Study*. National Commission on Teaching and America's Future. Retrieved from http://eric.ed.gov/?id=ED497176

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2008). *Who Leaves? Teacher Attrition and Student Achievement* (Working Paper No. 14022). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w14022

Boyd, D., H. Lankford, S. Loeb, J. Rockoff & J. Wyckoff. (2008). "The Narrowing Gap in New York City Teacher Qualifications and its Implications for Student Achievement in High-Poverty Schools" *Journal of Policy Analysis and Management*, 27(4), pp. 793-818.

Boyd, D., Lankford, H., Loeb, S., Ronfeldt, M., & Wyckoff, J. (2011). The role of teacher quality in retention and hiring: Using applications to transfer to uncover preferences of teachers and schools. *Journal of Policy Analysis and Management*, *30*(1), 88–110. http://doi.org/10.1002/pam.20545

Bryk, A. & Schneider, B. (2002). *Trust in Schools: A Core Resource for Improvement*. Russell Sage Foundation.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, *104*(9), 2633–79. http://doi.org/10.1257/aer.104.9.2633

Dee, T. S., & Wyckoff, J. (2015). Incentives, Selection, and Teacher Performance: Evidence from IMPACT. *Journal of Policy Analysis and Management*, *34*(2), 267–297. http://doi.org/10.1002/pam.21818

Goldhaber, D. (2015). Exploring the Potential of Value-Added Performance Measures to Affect the Quality of the Teacher Workforce. *Educational Researcher*, *44*(2), 87–95. http://doi.org/10.3102/0013189X15574905

Goldhaber, D. (2015). Teachers Clearly Matter, But Finding Effective Teacher Policies Has Proven Challenging. In H. Ladd and M. Geortz, (Eds). *Handbook of Research in Education Finance and Policy* (pp. 157-173). New York: Routledge, 2nd Edition.

Goldhaber, D., Gross, B., & Player, D. (2011). Teacher career paths, teacher quality, and persistence in the classroom: Are public schools keeping their best? *Journal of Policy Analysis and Management*, *30*(1), 57–87. http://doi.org/10.1002/pam.20549

Goldhaber, D., & Hansen, M. (2010). Using Performance on the Job to Inform Teacher Tenure Decisions. *The American Economic Review*, *100*(2), 250–255.

Goldhaber, D., Lavery, L., & Theobald, R. (2015). Uneven Playing Field? Assessing the Teacher Quality Gap Between Advantaged and Disadvantaged Students. *Educational Researcher 44*(5), 293-307.

Guin, K. (2004). Chronic Teacher Turnover in Urban Elementary Schools. *Education Policy Analysis Archives*, *12*(0), 42. http://doi.org/10.14507/epaa.v12n42.2004

Hanushek, E. A. (2009). Teacher Deselection. In D. Goldhaber & J. Hannaway (Eds.), *Creating a New Teaching Profession* (pp. 165–180). Washington, D.C.: Urban Institute Press.

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about Using Value-Added Measures of Teacher Quality. *The American Economic Review*, *100*(2), 267–271.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical Benchmarks for Interpreting Effect Sizes in Research. *Child Development Perspectives*, *2*(3), 172–177. http://doi.org/10.1111/j.1750-8606.2008.00061.x

Ingersoll, R. M., & Smith, T. M. (2003). The wrong solution to the teacher shortage. *Educational Leadership*, *60*(8), 30–33.

Isenberg, E., Max, J., Gleason, P., Potamites, L., Santillano, R., Hock, H., & Hansen, M. (2013). *Access to effective teaching for disadvantaged students* (No. NCEE 2014-4001). Washington, D.C.: National Center for Education Evaluation, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/pubs/20144001/pdf/20144001.pdf

Jackson, C. K. (2012). *Non-Cognitive Ability, Test Scores, and Teacher Quality: Evidence from 9th Grade Teachers in North Carolina* (Working Paper No. 18624). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w18624

Loeb, S., Miller, L. C., & Wyckoff, J. (2015). Performance Screens for School Improvement: The Case of Tenure Reform in New York City. *Educational Researcher*, 44(4), 199-212.

Papay, J., A. Bacher-Hicks, L. Page, W. Marinell, "The Challenge of Teacher Retention in Urban Schools: Evidence of Variation from a Cross-Site Analysis" Workking Paper, 2015.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, *73*(2), 417–458. http://doi.org/10.1111/j.1468-0262.2005.00584.x

Rockoff, J. E. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *The American Economic Review*, *94*(2), 247–252.

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can You Recognize an Effective Teacher when You Recruit One? *Education Finance and Policy*, *6*(1), 43-74.

Rockoff, J. E., & Speroni, C. (2010). Subjective and objective evaluations of teacher effectiveness. *The American Economic Review*, *100*(2), 261-266.

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How Teacher Turnover Harms Student Achievement. *American Educational Research Journal*, *50*(1), 4–36. http://doi.org/10.3102/0002831212463813

Rothstein, J. (2015). Teacher Quality Policy When Supply Matters. *American Economic Review*, *105*(1), 100–130. http://doi.org/10.1257/aer.20121242

Sass, T. R., Hannaway, J., Xu, Z., Figlio, D. N., & Feng, L. (2012). Value added of teachers in high-poverty schools and lower poverty schools. *Journal of Urban Economics*, *72*(2–3), 104–122. http://doi.org/10.1016/j.jue.2012.04.004

Smith, T. M., & Ingersoll, R. M. (2004). What Are the Effects of Induction and Mentoring on Beginning Teacher Turnover? *American Educational Research Journal*, *41*(3), 681–714. http://doi.org/10.3102/00028312041003681

Staiger, D. O., & Rockoff, J. E. (2010). Searching for Effective Teachers with Imperfect Information. *Journal of Economic Perspectives*, *24*(3), 97–118. http://doi.org/10.1257/jep.24.3.97

Thomsen, J. (2014). Teacher performance plays growing role in employment decisions. Education Commission of the States.

TNTP. (2012). The Irreplaceables: Understanding the Real Retention Crisis in America's Urban Schools. Brooklyn: TNTP.

Whitehurst, G., Chingos, M. & Lindquist, K. (2014). *Evaluating Teachers with Classroom Observations*. Brookings Institution.

Figures

Figure 1. Average IMPACT Scores of All General Education Teachers (Group 1 and 2) by Year
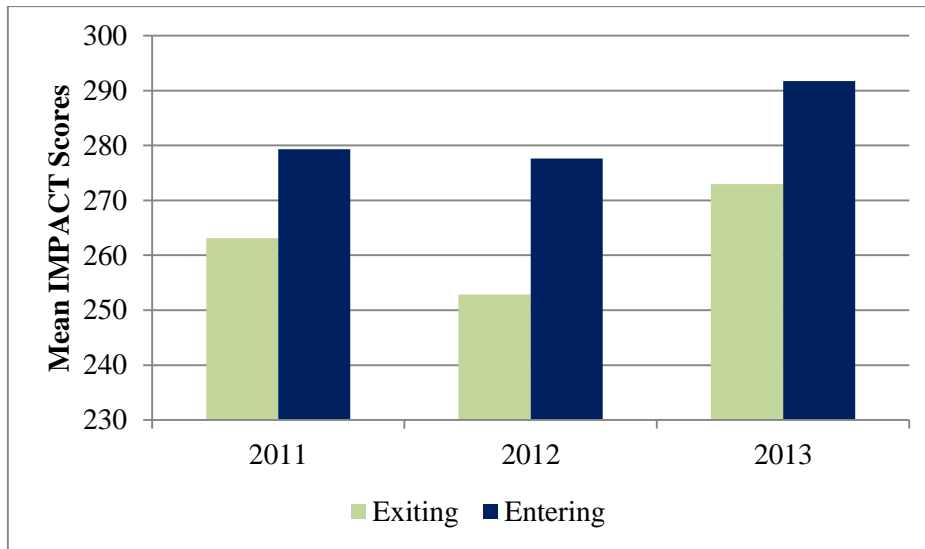
Notes. Results for 2011 indicate the average score for teachers who exited at the end of 2009-10 compared to those entering in 2010-11. Exits include teachers who retired, resigned or were terminated. Teachers leaving schools that closed are excluded.

Figure 2. Proportion of Teachers Exiting DCPS, by Teacher Performance and School Poverty
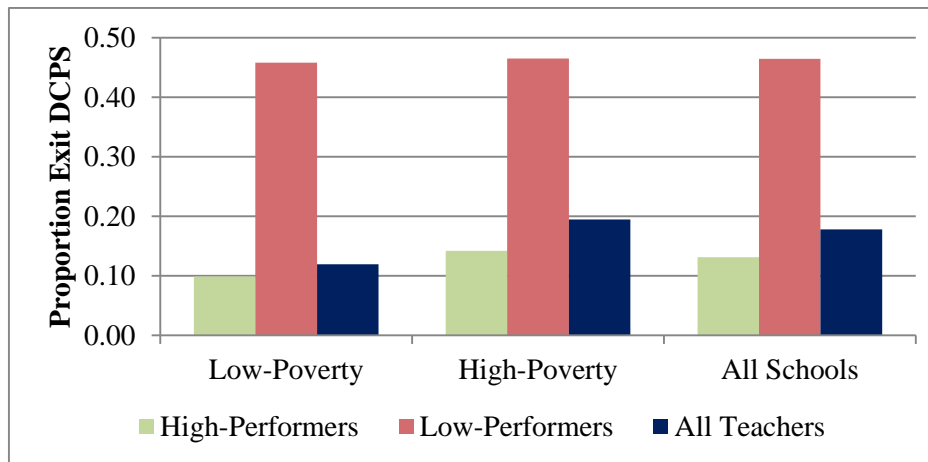
Notes. Teacher attrition indicates the average percent of teachers leaving DCPS at the end of 2009-10 through the end of 2011-12. Exits combine voluntary and involuntary exits, where voluntary exits include resignations and retirements and involuntary exits refer to teachers who were terminated due to performance. High-performers include teachers rated Effective of Highly Effective. Low-performers include teachers rated Ineffective or Minimally Effective.

Figure 3. Proportion of Exiting Teachers who are High- or Low-Performing, by School Poverty Status

Notes. Teacher attrition indicates the average percent of teachers leaving DCPS at the end of 2009-10 through the end of 2011-12. Exits combine voluntary and involuntary exits, where voluntary exits include resignations and retirements and involuntary exits refer to teachers who were terminated due to performance. High-performers include teachers rated Effective of Highly Effective. Low-performers include teachers rated Ineffective or Minimally Effective.

Figure 4. Average IMPACT Scores of All General Teachers (IMPACT Group 1 and Group 2) by Status of Exiting Teacher and Year
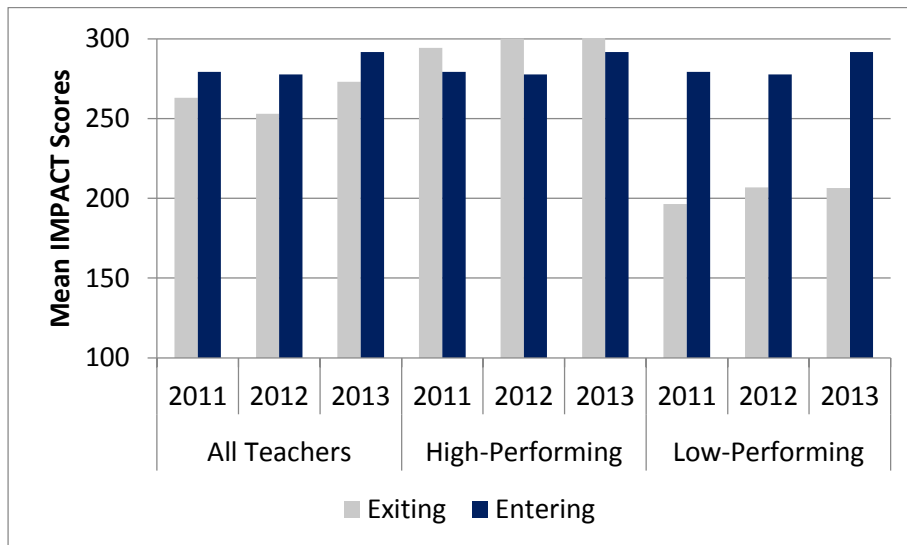
Notes. Results for 2011 indicate the average score for teachers who exited at the end of 2009-10 compared to those entering in 2010-11. Exiting scores are based on most recent IMPACT score. Scores of entering teachers are for all entering teachers as entering teachers cannot be linked to classroom of exiting teachers. Exits include teachers who retired, resigned or were terminated. Teachers leaving schools that closed are excluded.

Figure 5. Average IMPACT Scores of Teachers Who Are Matched to Students with Math Achievement Scores (IMPACT Group 1) by Year
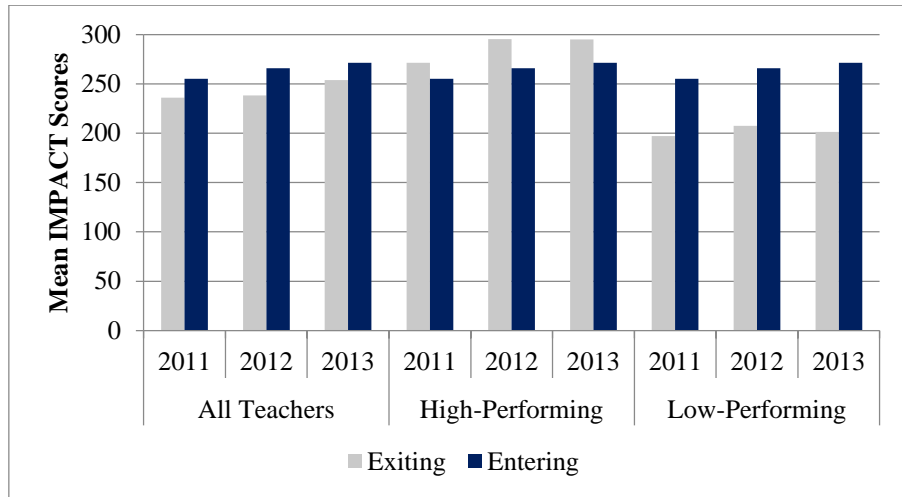
Notes. Results for 2011 indicate the average score for teachers who exited at the end of 2009-10 compared to those entering in 2010-11. Exiting scores are based on most recent IMPACT score. Scores of entering teachers are for all entering teachers as entering teachers cannot be linked to classroom of exiting teachers. Exits include teachers who retired, resigned or were terminated. Teachers leaving schools that closed are excluded.

Figure 6. Average Individual Value-Added Scores of Teachers Who Are Matched to Students with Math Achievement Scores (IMPACT Group 1) by Status of Exiting Teacher and Year
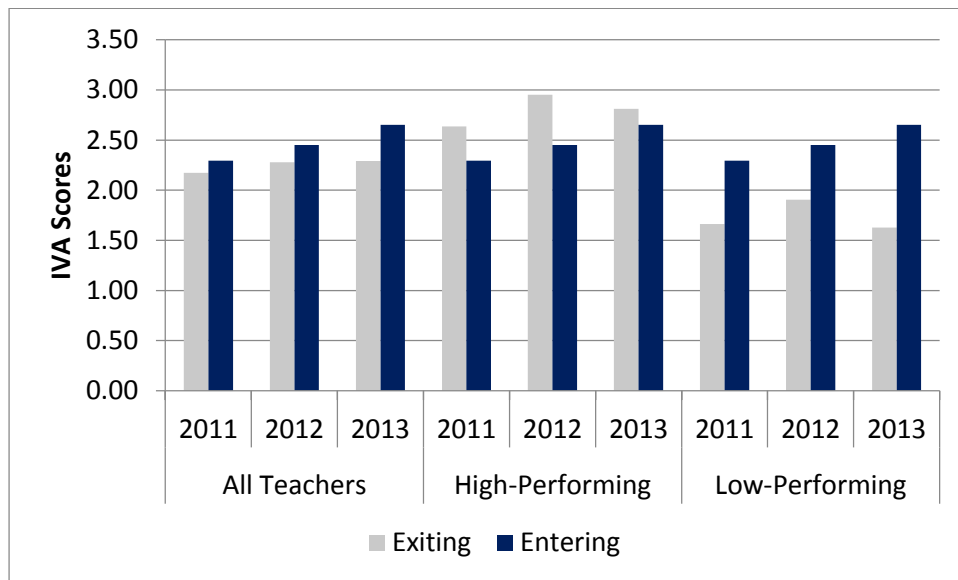
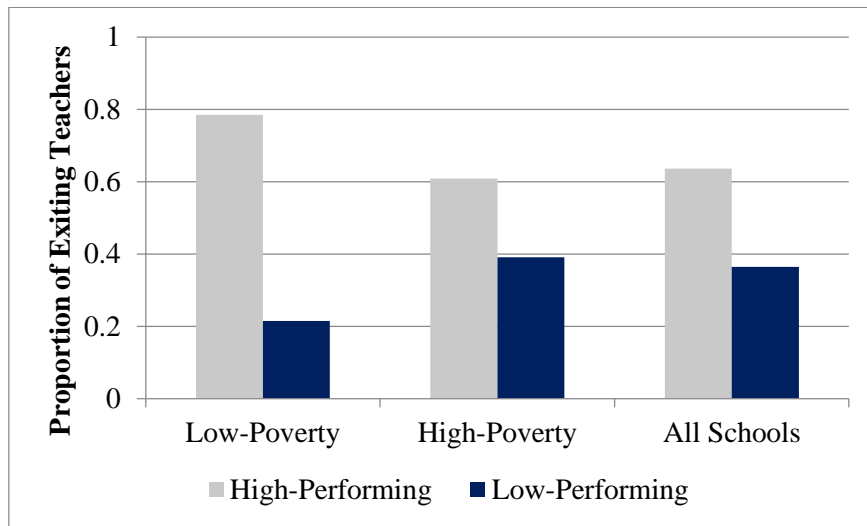*Notes*. Results for 2011 indicate the average score for teachers who exited at the end of 2009-10 compared to those entering in 2010-11. Exiting scores are based on most recent IMPACT score. Scores of entering teachers are for all entering teachers as entering teachers cannot be linked to classroom of exiting teachers. Exits include teachers who retired, resigned or were terminated. Teachers leaving schools that closed are excluded.

Tables

Table 2. Descriptive Statistics for Math and Reading, Various Samples, 2011-2013

|	Math Samples				Reading Sample			
	Unrestricted	Base	Unbalanced	Balanced	Unrestricted	Base	Unbalanced	Balanced
Average Student Characteristics (N=56,564 student-year observations)								
Students per s-g-y cell	50.6	51.2	51.0	51.1	50.6	51.4	51.7	52.8
	(44.6)	(45.1)	(44.7)	(44.9)	(44.6)	(45.0)	(45.5)	(46.7)
Proportion Male	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
Proportion Black	0.77	0.77	0.77	0.76	0.77	0.77	0.77	0.76
	(0.28)	(0.29)	(0.29)	(0.29)	(0.28)	(0.28)	(0.29)	(0.29)
Proportion Hispanic	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
	(0.21)	(0.21)	(0.21)	(0.22)	(0.21)	(0.21)	(0.21)	(0.21)
Proportion LEP	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
Proportion SpEd	0.18	0.18	0.18	0.17	0.18	0.18	0.18	0.17
	(0.10)	(0.09)	(0.09)	(0.09)	(0.10)	(0.09)	(0.09)	(0.09)
Proportion FRPL	0.69	0.70	0.70	0.69	0.69	0.70	0.70	0.70
	(0.23)	(0.23)	(0.23)	(0.23)	(0.23)	(0.23)	(0.23)	(0.23)
Residualized achievement	-0.02	-0.02	-0.02	-0.01	-0.03	-0.03	-0.03	-0.03
	(0.23)	(0.23)	(0.23)	(0.23)	(0.18)	(0.18)	(0.18)	(0.18)
Average Teacher Characteristics (N=1,873 teacher-year observations)								
Teachers per s-g-y cell	1.68	1.68	1.68	1.70	1.86	1.86	1.86	1.89
	(0.86)	(0.87)	(0.87)	(0.88)	(0.93)	(0.93)	(0.93)	(0.95)
Any Exit	0.21	0.21	0.20	0.19	0.19	0.20	0.19	0.19
	(0.36)	(0.36)	(0.35)	(0.34)	(0.34)	(0.35)	(0.34)	(0.34)
High-Performer Exit	0.10	0.10	0.10	0.10	0.10	0.10	0.09	0.09
	(0.26)	(0.27)	(0.26)	(0.25)	(0.26)	(0.26)	(0.25)	(0.25)
Low-Performer Exit	0.11	0.10	0.10	0.10	0.10	0.10	0.10	0.10
	(0.27)	(0.27)	(0.26)	(0.26)	(0.26)	(0.26)	(0.26)	(0.26)
IMPACT score	283.5	283.7	283.7	286.3	284.8	285.2	285.2	286.6
	(51.0)	(51.2)	(51.2)	(50.7)	(48.7)	(48.7)	(48.7)	(48.4)
Teaching experience	9.55	9.63	9.63	9.91	9.30	9.37	9.37	9.56
	(6.92)	(6.89)	(6.89)	(6.92)	(6.77)	(6.76)	(6.76)	(6.72)
Average School Characteristics								
Number of unique schools	100	97	97	88	100	97	97	90
% High Poverty	80	80.41	80.41	79.55	80	80.41	80.41	81.11
% Elementary	64	64.95	64.95	67.05	64	64.95	64.95	65.56
% Middle	14	13.40	13.40	12.50	14	13.40	13.40	14.44
% Senior High School	1	1.03	1.03	1.14	1	1.03	1.03	1.11
% Education Campus	20	20.62	20.62	19.32	20	20.62	20.62	18.89
School-Grade-Year Obs.	838	751	734	663	838	753	733	666

Notes. Unrestricted sample includes school-grade-year cells which contain non-missing data for all variables in our model. The Base sample restricts the sample to school-grade-year cells which contain non-missing outcome data in two consecutive years (to form the first differences). The Unbalanced sample further restricts to school-grade-year cells which contain both IMPACT scores and student achievement. The Balanced sample is limited to school-grade cells which contain all three years of first differences.

Table 3. Effect of teacher turnover on IMPACT scores and math and reading student achievement

	Math				Reading			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	IMPACT score	DCCAS	IMPACT score	DCCAS	IMPACT score	DCCAS	IMPACT score	DCCAS
All Exits	17.359*	0.079**			15.066*	0.046~		
	(6.973)	(0.03)			(6.244)	(0.024)		
High-Performers			-29.720**	-0.055			-17.798*	-0.047
			(8.486)	(0.039)			(7.697)	(0.034)
Low-performers			63.838**	0.210**			46.129**	0.136**
			(8.071)	(0.041)			(7.987)	(0.03)
Student Controls		X		X		X		X
Observations	663	663	663	663	666	666	666	666
R-squared	0.035	0.015	0.138	0.045	0.035	0.017	0.087	0.04

Notes. Robust standard errors reported in parentheses. Asterisks indicate statistical significance (~ p< .10, *p < 0.05, ** p < 0.01). All models include year fixed effects and controls for teacher movement within and across schools. Student controls account for the year-to-year, across cohort change in the percent of students in a school-grade-year cell who are Black, Hispanic, other non-White race/ethnicity, limited-English proficient, special education, or FRPL-eligible.

Table 4. Effect of teacher turnover on IMPACT scores and math and reading student achievement by school poverty status

	Math				Reading			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	IMPACT score	DCCAS	IMPACT score	DCCAS	IMPACT score	DCCAS	IMPACT score	DCCAS
All Exits								
Low-Poverty	21.714	0.006			1.738	-0.038		
	(19.301)	(0.082)			(8.727)	(0.043)		
High-Poverty	16.793*	0.084**			16.032*	0.052*		
	(7.259)	(0.03)			(6.548)	(0.025)		
High-Performer Exits								
Low-Poverty			23.648	-0.004			1.922	-0.041
			(21.962)	(0.097)			(9.892)	(0.05)
High-Poverty			-39.234**	-0.064			-20.437~	-0.048
			(-8.596)	(0.042)			(8.575)	(0.038)
Low-Performer Exits								
Low-Poverty			NA	NA			NA	NA
High-Poverty			64.075**	0.209**			46.761**	0.138**
			(8.171)	(0.041)			(8.104)	(0.03)
Student Controls		X		X		X		X
Observations	663	663	663	663	666	666	666	666

Notes. Robust standard errors reported in parentheses. Asterisks indicate statistical significance (~ $p<0.10$, *$p < 0.05$, ** $p < 0.01$). All models include year fixed effects and controls for teacher movement within and across schools. Student controls account for the year-to-year, across cohort change in the percent of students in a school-grade-year cell who are Black, Hispanic, other non-White race/ethnicity, limited-English proficient, special education, or FRPL-eligible. We do not include estimates for low-performer exits in low-poverty schools as these are found in only three school-grade cells.

Table 5. Effect of teacher turnover on IMPACT scores and math and reading student achievement, by year

	Math				Reading			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	IMPACT score	DCCAS	IMPACT score	DCCAS	IMPACT score	DCCAS	IMPACT score	DCCAS
All Exits								
2011	13.053	0.092			2.113	-0.039		
	(11.876)	(0.061)			(9.113)	(0.043)		
2012	20.706	0.025			23.396**	0.046		
	(14.355)	(0.059)			(10.354)	(0.039)		
2013	20.017~	0.112**			20.258~	0.105*		
	(11.429)	(0.041)			(10.895)	(0.041)		
High-Performer Exits								
2011			-15.404	-0.022			-38.553**	-0.161**
			(13.793)	(0.071)			(9.513)	(0.054)
2012			-53.682**	-0.277**			-9.302	-0.042
			(15.312)	(0.059)			(12.900)	(0.049)
2013			-21.426	0.057			-9.969	0.008
			(13.399)	(0.057)			(11.766)	(0.059)
Low-Performer Exits								
2011			43.824**	0.215*			31.473**	0.050
			(14.939)	(0.100)			(10.267)	(0.049)
2012			74.931**	0.243**			54.623**	0.133**
			(16.914)	(0.067)			(12.376)	(0.050)
2013			70.166**	0.179**			53.750**	0.208**
			(12.678)	(0.051)			(15.822)	(0.047)
Student Controls		X		X		X		X
Observations	663	663	663	663	666	666	666	666

Notes. Robust standard errors in parentheses. Asterisks indicate statistical significance (~ p<0.10, *p < 0.05, ** p < 0.01). All models include controls for teacher movement within and across schools. Student controls account for the across cohort change in the percent of students in a school-grade-year cell who are Black, Hispanic, other non-White race/ethnicity, LEP, special education, or FRPL-eligible

Technical Appendix

Our analysis combines student data, teacher data, student-teacher linkage rosters, and school data.

Student Data. The primary outcome of interest (i.e., the year-to-year change in residualized and standardized student achievement at the school-grade level) is constructed using student-level achievement data from DCPS' Comprehensive Assessment System (DC-CAS). DCPS piloted the DC-CAS in the spring of the 2005-06 school year and has since then used the same assessment system to evaluate annual math and English Language Arts (ELA) achievement for third through eighth grade students. For this analysis we employ student-level achievement data from 2008-09 through 2012-13 that identifies each student's assigned school, grade, test grade, test subject, raw and scale score, proficiency level, gender, race, free/reduced price lunch eligibility, special education status, and limited English proficiency status. Scale scores were standardized by grade, subject, and year, excluding students who were tested in a grade other than their assigned grade.[17] We then used students' standardized scale scores to create a residualized measure of student achievement that accounts for each students' prior achievement and demographic characteristics. Specifically, we recover residuals from the following linear regression:

$$A_{it} = \alpha_0 + \beta_1 A_{it-1} + \beta_2 Male_i + \beta_3 Black_i + \beta_4 Hispanic_i + \beta_5 OtherRace_i +$$

$$\beta_6 LEP_{it} + \beta_7 SpEd_{it} + \beta_8 FRPL_{it} + \varepsilon_{it}$$

where A_{it} is a student's standardized achievement in math or reading at time t, A_{it-1} is the students' lagged standardized achievement in the same subject, and the remaining predictors are indicator variables identifying observable student characteristics. We only create residualized achievement for students who progress normally through the grade sequence. Moreover, because we control for prior achievement, we are only able to create residualized achievement for fourth through eighth grade students. After creating these analytic variables, we collapse the student-level achievement data to the school-grade-year level. Each row in this collapsed file is a school-grade-year cell. For each school-grade-year cell, collapsed variables identify the number of students, average residualized achievement, the percent of students by race, the percent of students identified as having

---

[17] This affects 0.22 percent of student-year observations.

limited English proficiency (LEP), and the percent of students eligible for special education (SpEd) and free/reduced price lunch (FRPL).

Teacher Data

We combine teacher-level IMPACT data with administrative data to identify the prevalence and character of teacher turnover at the end of each school year. IMPACT data from 2009-10 through 2012-13 informs us of teachers' final IMPACT score and IMPACT rating in each year. Teacher-level administrative data over the same time period links each teacher to a school, allowing us to track teacher movement within and out of DCPS. We identify teachers as having left DCPS if they are no longer assigned to a school from one year to the next. We exclude teachers leaving schools that closed to avoid conflating the effect of teacher turnover with the effect of school closure.[18] We use IMPACT ratings to divide teacher exits into two sub-categories: teacher exits are considered low-performer exits if the teacher was rated Ineffective or Minimally Effective in their last year with DCPS; exiting teachers rated Effective or Highly Effective in their last year are considered high-performer exits. We also identify teachers who move to a new position in the same school and teachers who move to a new school in the district to control for teacher movement out of a school-grade-year cell that does not result in the teacher leaving DCPS.

Student-Teacher Linkage Rosters

Teacher-level administrative data match teachers to schools but do not indicate the grade of instruction. In order to link teachers to grades, we rely on linking rosters that match teachers to students in tested grades and subjects. As previously described, student-level achievement data identify each students' assigned grade and school. Thus, teachers are assigned to school-grade-year cells based on the students to whom they are linked and the students' school and grade assignment. We allow student-teacher links to vary by subject and create separate analytic samples for each subject. In both math and ELA teachers can be linked to students in different grades in

---

[18] Teacher exit status is missing for teachers who leave schools that closed. This only affects eleven teacher-school-grade-year records.

the same school, but students are only linked to one school and grade in a given year. We connect teachers to school-grade-year cells if they are linked to at least ten students in a given year.

We collapse the data to the school-grade-year level and identify the proportion of students in a school-grade-year cell who experience each type of teacher turnover in the prior school year. For each school-grade-year cell, variables identify the number of teachers, the average IMPACT score, and the proportion of students who experienced each type of teacher turnover in the prior year. When we merge this dataset with the student achievement data described above, approximately 85 percent of the school-grade-year cells from the collapsed student data merged with the collapsed teacher data.

School Data

Four schools were dropped due to missing data for school poverty status for these schools; resulting in missing data in 66 school-grade-year cells. School-level data provide three important indicators: school poverty status, school type, and school closure status. DCPS categorizes schools as either high- or low-poverty based on the percent of students eligible for free or reduced price lunch (FRPL) in the current school year. Because both the percent of students eligible for FRPL and the high-poverty assignment threshold have changed over time, poverty status can also vary over time for a given school.[19] School type identifies elementary, middle, and high schools. In addition, DCPS identifies some schools as Education Campuses. These are schools that span elementary and middle school grades. Other school types include Special Education Schools and Program Schools. Program Schools are schools that serve unique student populations. For instance, one Program School is located within a youth detention center. Program Schools and Special Education Schools are not included in our sample because students in these schools do not participate in the regular DC CAS. During our sample period, 27 schools were closed. Thirteen school closures occurred at the end of 2012-13, so these school closures do not affect our sample. Of the 14 school closures that occurred prior to 2012-13, only six occurred in traditional schools that would be included in our analytic sample.

---

[19] Prior to 2012-13, schools were identified as high-poverty if more than 60 percent of students were eligible for FRPL. In 2012-13, the threshold was revised down to 50 percent, causing more schools to be identified as high-poverty.

Stacked First-Differenced Sample

We create variables that capture the year-to-year change in each outcome and predictor variable at the school-grade-year level. Outcome variables include the change in average IMPACT scores in each school-grade-year cell and the change in average residualized student achievement in each school-grade-year cell. Predictor variables include changes in student demographics in each school-grade-year cell. The predictor of interest, the proportion of students in each school-grade-year cell that experienced teacher turnover in the prior year, is not differenced. In this manner, the predictor variable captures the within school-grade cell change in residualized student achievement that is associated with the proportion of students that experienced teacher turnover in the prior year.

Appendix Tables

Appendix Table 1. Differences in means of variables in the Unrestricted and Balanced samples

	Math			Reading		
	Dropped	Retained		Dropped	Retained	
Average IMPACT score	264.63	286.26	**	272.35	286.56	**
	(4.98)	(1.97)		(5.13)	(1.88)	
	N = 98			N = 92		
Average Residualized Achievement	-0.09	-0.01	**	-0.03	-0.02	
	(0.02)	(0.01)		(0.02)	(0.01)	
	N = 116			N = 113		
Any Exit (dosage)	0.28	0.19	*	0.19	0.19	
	(0.04)	(0.01)		(0.03)	(0.01)	
	N = 113			N = 112		
Proportion FRPL	0.69	0.69		0.67	0.70	
	(0.02)	(0.01)		(0.02)	(0.01)	
	N = 128			N = 125		
Proportion SpEd	0.21	0.17	**	0.18	0.17	
	(0.01)	(0.00)		(0.01)	(0.00)	
	N = 128			N = 125		
Proportion LEP	0.07	0.08		0.06	0.08	*
	(0.01)	(0.00)		(0.01)	(0.00)	
	N = 128			N = 125		
Proportion Black	0.86	0.76	**	0.85	0.76	**
	(0.02)	(0.01)		(0.02)	(0.01)	
	N = 128			N = 125		
Proportion Hispanic	0.10	0.15	*	0.10	0.15	*
	(0.02)	(0.01)		(0.02)	(0.01)	
	N = 128			N = 125		
Proportion High-poverty	0.91	0.82	**	0.87	0.83	
	(0.02)	(0.01)		(0.03)	(0.01)	
	N = 165			N = 162		
Observations	175	663		172	666	

Notes. Asterisks indicate a statistically significant difference between dropped and retained cells based on a two-tailed t-test with equal variance (~ p<0.10, *p < 0.05, ** p <0 .01). N indicates the number of dropped cells that contain data. "Any Exit (dosage)" "Proportion FRPL" "Proportion SpEd" and "Proportion LEP" are the proportion of students who: experiencing teacher turnover, are eligible for free/reduced price lunch, receive special education services and are identified as limited-English proficient, respectively.

Appendix Table 2. Effect of teacher turnover on math student achievement by sample*

	Base Sample		Unbalanced Sample		Balanced Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
All Exits	0.050~		0.070*		0.079**	
	(0.028)		(0.028)		(0.030)	
High-Performer Exits		-0.073*		-0.060~		-0.055
		(0.036)		(0.036)		(0.039)
Low-Performer Exits		0.171**		0.207**		0.210**
		(0.038)		(0.039)		(0.041)
School Fixed Effects						
Observations	751	751	734	734	663	663

*Notes: Robust standard errors in parentheses ~ p<0.10, * p<0.05, ** p<0.01. Estimates including school fixed effects are very similar with the same pattern of statistical significance. Results available from authors.

Appendix Table 3. Effect of teacher turnover on reading student achievement by sample*

	Base Sample		Unbalanced Sample		Balanced Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
All Exits	0.026		0.039~		0.046~	
	-0.023		-0.023		-0.024	
High-Performer Exits		-0.066*		-0.056~		-0.047
		-0.032		-0.033		-0.034
Low-Performer Exits		0.117**		0.125**		0.136**
		-0.028		-0.028		-0.03
School Fixed Effects						
Observations	753	753	733	733	666	666

*Notes: Robust standard errors in parentheses ~ p<0.10, * p<0.05, ** p<0.01. Estimates including school fixed effects are very similar with the same pattern of statistical significance. Results available from authors.

Appendix Table 4. Effect of teacher turnover on changes in student demographics, balanced math sample*

	Δ Black		Δ Hispanic		Δ Other Race		Δ FRPL		Δ SpEd		Δ LEP	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
All Exits	0.001		0.004		-0.003		0.024~		0.002		-0.013	
	(0.009)		(0.009)		(0.003)		(0.014)		(0.010)		(0.008)	
High-Performer Exits		-0.010		0.009		-0.002		0.020		-0.013		-0.024*
		(0.013)		(0.013)		(0.004)		(0.018)		(0.013)		(0.011)
Low-Performer Exits		0.012		0.000		-0.003		0.027		0.016		-0.002
		(0.010)		(0.010)		(0.005)		(0.017)		(0.014)		(0.011)
School Fixed Effects												
Observations	663	663	663	663	663	663	663	663	663	663	663	663
R-squared	0.004	0.007	0.003	0.004	0.019	0.019	0.192	0.192	0.005	0.008	0.045	0.050

*Notes: Robust standard errors in parentheses ~ p<0.10, ** p<0.01, * p<0.05. In estimates that include the school fixed none of the variables are statistically significant at tradition levels. Similar analyses for reading produce no statistically significant estimates.

Appendix Table 5. Effect of including school fixed effects on teacher turnover estimates on math student achievement

	No FE (1)	School FE (2)	School-Year FE (3)	No FE (4)	School FE (5)	School-Year FE (6)
All Exits	0.079**	0.058	0.009			
	(0.030)	(0.038)	(0.035)			
High-Performers				-0.055	-0.083	-0.080~
				(0.039)	(0.050)	(0.041)
Low-performers				0.210**	0.199**	0.100*
				(0.041)	(0.054)	(0.049)
Student Controls	X	X	X	X	X	X
School Fixed Effects		X			X	
School-Year Fixed Effects			X			X
Observations	663	663	663	663	663	663

Notes. Robust standard errors reported in parentheses. Asterisks indicate statistical significance (~ p<.10, *p < 0.05, ** p < 0.01). All models include year fixed effects and controls for teacher movement within and across schools. Student controls account for the year-to-year, across cohort change in the percent of students in a school-grade-year cell who are Black, Hispanic, other non-White race/ethnicity, limited-English proficient, special education, or FRPL-eligible.

Appendix Table 6. Effect of including school fixed effects on teacher turnover estimates on reading student achievement

	No FE	School FE	School-Year FE	No FE	School FE	School-Year FE
	(1)	(2)	(3)	(4)	(5)	(6)
All Exits	0.046~	0.038	0.017			
	(0.024)	(0.032)	(0.030)			
High-Performers				-0.047	-0.057	-0.074
				(0.034)	(0.042)	(0.051)
Low-performers				0.136**	0.133**	0.105*
				(0.030)	(0.041)	(0.044)
Student Controls	X	X	X	X	X	X
School Fixed Effects		X			X	
School-Year Fixed Effects			X			X
Observations	666	666	666	666	666	666

Notes. Robust standard errors reported in parentheses. Asterisks indicate statistical significance (~ p<.10, *p < 0.05, ** p < 0.01). All models include year fixed effects and controls for teacher movement within and across schools. Student controls account for the year-to-year, across cohort change in the percent of students in a school-grade-year cell who are Black, Hispanic, other non-White race/ethnicity, limited-English proficient, special education, or FRPL-eligible.

Appendix Table 7. Effects of Teacher Turnover in 2012 on the Change in Student Achievement from 2009-10 to 2010-11

	Math		Reading	
	(1)	(2)	(3)	(4)
	DCCAS	DCCAS	DCCAS	DCCAS
All Exits	0.070		0.015	
	(0.062)		(0.050)	
High-Performers		0.103		0.032
		(0.081)		(0.061)
Low-performers		0.033		-0.004
		(0.080)		(0.073)
Student Controls	X	X	X	X
Observations	663	663	666	666