AERA Open
April-June 2020, Vol. 6, No. 2, pp. 1–21
DOI: 10.1177/2332858420932235
Article reuse guidelines: sagepub.com/journals-permissions
© The Author(s) 2020. http://journals.sagepub.com/home/ero

Teacher Evaluation and Teacher Turnover in Equilibrium: Evidence From DC Public Schools

Jessalynn James
Annenberg Institute, Brown University

James H. Wyckoff
University of Virginia

Teacher turnover is an enduring concern in education policy and can incur substantial costs to students. Policies often address turnover broadly, yet effects turn on net differences in the effectiveness of exiting and entering teachers, in addition to the disruption dealt to classrooms. Recent research has shown mixed effects of teacher evaluation policies, but even where evaluation-induced differential turnover initially benefited students, gains might disappear or reverse as the stock of less effective teachers exits and if more effective teachers view high-stakes evaluation as burdensome. We examine evaluation–induced changes to the composition of exiting and entering teachers in Washington, D.C., the net effect of turnover on student achievement, and the role that evaluation played in teacher turnover. We find that turnover continues to improve teaching skills and student achievement, although effects have diminished. We find little evidence that high-performing teachers' exit is associated with the evaluation system.

Keywords: *descriptive analysis, econometric analysis, educational policy, evaluation, quasi-experimental analysis, school/teacher effectiveness, teacher turnover, urban education*

Few topics in education policy receive more attention than teacher turnover. Research documents substantial negative effects on student achievement. Effects are felt disproportionately by schools with more low-performing and Black students (e.g., Ronfeldt et al., 2013)—the students for whom teacher turnover is greatest and for whom receiving an effective replacement is least likely. But turnover is not a concern for low-performing schools alone, nor is it a recent phenomenon. Teacher turnover arises in policy discussions spanning teacher preparation, school finance, student achievement, accountability, and school leadership. The scope of its implications has created a sense of urgency, and policymakers embrace a variety of proposals to mitigate teacher turnover. These policies, however, along with much research, often treat teacher retention as unambiguously beneficial (Carver-Thomas & Darling-Hammond, 2019; Ingersoll, 2001; National Commission on Teaching and America's Future, 2003), despite broad recognition that teaching quality varies widely. In the absence of valid and reliable measures of teacher effectiveness and given the available evidence that teacher turnover harms students, such an approach may be appropriate. However, most states and school districts have recently revised their teacher evaluation systems, providing opportunities for more targeted—and therefore more effective—turnover policies.

Turnover's effects on student outcomes depend on two mechanisms: (1) changes in the composition of teaching effectiveness and (2) the disruptive effect on teachers who remain and their students (Ronfeldt et al., 2013). The compositional effect is conceptually uncertain and turns on the differential effectiveness of exiting and entering teachers. A decade ago, The New Teacher Project (TNTP) highlighted this issue in *The Widget Effect* (Weisberg et al., 2009), documenting the widespread practice of treating teachers, regardless of effectiveness, as interchangeable. The report proposed that valid and reliable teacher evaluation would provide credible evidence of strong performance by some teachers. This recognition would be both an intrinsic reward and a mechanism for incentivizing performance and retention through compensation and advancement opportunities. Conversely, the report hypothesized that teachers who received poor evaluations would be more likely to *voluntarily* exit when presented with credible evidence of their weaknesses.

The potential for teacher evaluation to enable more effective teacher retention policies rests on several assumptions. First, evaluation systems implemented at scale must be

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

reliable and valid—for example, accepted by teachers as credible evidence of their skills. Second, teachers identified as poor performers must be replaced by teachers who are at least as effective. Third, high-performing teachers must not find the stress of ongoing high-stakes evaluation so burdensome or threatening to their job security that it diminishes the supply of high-quality applicants to the district or produces a net increase in their turnover. Finally, even if teacher evaluation facilitates improvements to teaching and student achievement, these gains may not be sustainable as the stock of relatively ineffective teachers is reduced, if the quality of applicants changes in response to the system's incentives, and if administrators' focus wanes in the face of resistance to evaluation and other demands.

In this article, we explore these issues in the context of the District of Columbia Public Schools (DCPS). We focus on three research questions:

- **Research Question 1:** What is the effectiveness of exiting and entering DCPS teachers, and how has that changed over the course of the implementation of its high-stakes evaluation system?
- **Research Question 2:** What effect does teacher turnover have on teaching skills and student achievement in DCPS several years after the evaluation reform's implementation, and does this differ by teaching effectiveness?
- **Research Question 3:** To what extent do high- and low-performing teachers cite the evaluation system as a reason for their decision to exit, and how has that changed over time?

DCPS is a particularly appealing place to explore these questions. For a decade, DCPS has employed one of the most rigorous teacher evaluation systems in the United States. DCPS also experiences relatively high turnover—nearly 20% of teachers exit DCPS each year and 5-year attrition is 57% (Figure 1)—raising concerns about the effect on students. There are also concerns that IMPACT, DCPS's teacher evaluation system, might play a role in such turnover (e.g., Levy, 2018).

Background

A large literature examines various aspects of teacher turnover,[1] much of which focuses on overall turnover in specific school districts or states, factors associated with turnover, and policies that may influence turnover. In these analyses, exiting teachers are typically treated as homogenous, without consideration for whether they can be replaced by more effective teachers, despite long-standing recognition that strategic retention can be a powerful lever for improving teacher effectiveness (e.g., Smith & Handler, 1979; TNTP, 2012).

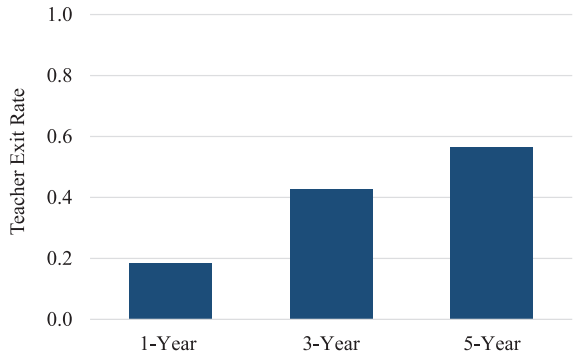

FIGURE 1 *Proportion of teachers exiting DCPS over 1, 3, and 5 years, 2012–2013 through 2016–2017.*
Note. For all years except 2017, a teacher is retained if they taught and received an IMPACT rating in t and were a classroom teacher with an IMPACT rating in $t + 1$ or $t + 2$. For 2017 teachers, retention includes those who were a classroom teacher with an IMPACT rating in $t + 1$ (2018). DCPS = District of Columbia Public Schools.

Growing evidence, however, finds that less effective teachers are more likely to exit than their higher performing peers (Boyd et al., 2008; Feng & Sass, 2017; Goldhaber et al., 2010; Hanushek et al., 2005; Murnane, 1984; Papay et al., 2017). The mechanisms by which this occurs are unclear. Smaller scale pilots found that information on effectiveness led to increased turnover among low-performing teachers but with no discernable effect on student achievement (Loeb et al., 2015; Rockoff et al., 2012; Sartain & Steinberg, 2016). The same patterns are evidenced, however, in settings where teachers or their supervisors do not have systematic access to performance information (Boyd et al., 2008; Hanushek et al., 2005; Papay et al., 2017).

Nonetheless, the composition effect of teacher turnover remains uncertain, as it depends on the differential effects of exiting *and entering* teachers. The literature on teacher hiring raises important questions about whether hiring officials identify applicants who will become effective teachers (e.g., Jacob et al., 2018), while in practice some principals appear differentially able to retain more effective teachers (Cohen et al., 2020; Grissom & Bartanen, 2019). Little evidence bears directly on the relationship between teacher evaluation, teacher turnover, and student achievement. A couple of notable exceptions provide a foundation for our research, demonstrating that teacher evaluation can induce less effective teachers to disproportionately exit, raising the quality of teacher effectiveness (Cullen et al., 2019; Dee et al., 2019; Dee & Wyckoff, 2015; Sartain & Steinberg, 2016; Stecher et al., 2018). In a telling example, teacher evaluation reform in Houston increased low-performing teachers' voluntary exit by 6.2% (Cullen et al., 2019). When the average low-performing teacher exited, student achievement improved by about 13% of a standard deviation (*SD*). However, because low-performing teachers represent a small share of reform-induced exits (4.3%), the overall effectiveness of Houston

teachers was not meaningfully altered. Similarly, teacher evaluation, as implemented in the school districts and charter school management organizations participating in the Gates-funded Intensive Partnerships for Effective Teaching, increased ineffective teachers' attrition but not enough to improve student achievement (Stecher et al., 2018).

Conversely, many policies have employed performance-based financial rewards to incentivize high-performing or highly qualified teachers' retention, with some—but not always lasting—success (Clotfelter et al., 2008; Cowan & Goldhaber, 2018; Glazerman et al., 2013; Glazerman & Seifullah, 2012; Springer et al., 2016). The fidelity of evaluation implementation in some of these settings (e.g., Cullen et al., 2019; Stecher et al., 2018) raises concerns, but these examples provide a cautionary note regarding teacher evaluation's potential to broadly improve student achievement. The experience in DCPS is different and worth a closer look.

Teacher Evaluation in Washington, D.C.

Ten years ago, DCPS introduced a bundle of reforms to address a long history of dysfunction and low student performance. Central to these reforms was a rigorous teacher evaluation system—IMPACT (National Research Council, 2015), which had three main design components:

- Every teacher was assessed yearly using multiple measures of teaching effectiveness, including five standards-based classroom observations conducted by calibrated observers and some measure of student achievement.
- Teachers received professional development supports in the form of feedback following each formal classroom observation.
- Teachers who scored in the lowest rating category (Ineffective) were subject to immediate dismissal. Low-performing teachers (Minimally Effective) were also subject to dismissal if they failed to improve. Teachers rated Highly Effective were eligible for large financial rewards (bonuses and base pay increases) and professional opportunities.

Unlike many evaluation systems, DCPS devoted substantial resources to the rigorous implementation of IMPACT, incorporating many best practices then emerging from research (Toch, 2018). For example, while in most systems nearly all teachers are identified as effective or better (Kraft & Gilmour, 2017), DCPS differentiated performance. During its first 3 years, IMPACT identified 15% of teachers as less than Effective (Ineffective or Minimally Effective) and 15% as Highly Effective.

IMPACT was controversial from the outset, with legitimate concerns about the assessment's fairness, whether teachers would receive the feedback and support necessary to improve, and whether the stress of teaching in such a high-stakes environment would drive high-performing teachers to neighboring districts.[2] Nevertheless, the new program demonstrated success at achieving some of its primary goals. Analyses of IMPACT's first 3 years found that low-performing teachers subject to IMPACT's strongest incentives experienced a large increase in voluntary turnover, and those who chose to remain improved their performance (Dee & Wyckoff, 2015). Specifically, using a regression discontinuity (RD) design, Dee and Wyckoff (2015) found that teachers in 2010–2011 and 2011–2012 who were rated Minimally Effective (and therefore had a year to improve or face dismissal) but were near the threshold for Effective were 50% more likely to voluntarily exit. Those who remained improved their performance by 27% of an *SD* relative to otherwise-similar teachers not facing this incentive. IMPACT's generous financial incentives for Highly Effective teachers were estimated to have positive but statistically insignificant effects on retention, but these financial rewards induced teachers receiving their first Highly Effective rating (they would receive a large base pay increase upon a second Highly Effective rating) to improve their IMPACT scores by 24% of an *SD*. These results demonstrate the potential of high-stakes teacher evaluation to induce low-performing teachers' voluntary turnover or improvement and already high-performing teachers' improvement. *Importantly, these results are specific to the teachers near these ratings thresholds and say little about the vast majority of teachers for whom these incentives don't apply*, nor do they necessarily translate into improved student outcomes.

Adnot et al. (2017) provided more insight into the latter in their examination of the effect of teacher turnover on teaching quality and student achievement in DCPS during the same period. They found the following:

- On average, turnover in DCPS improved teaching quality (0.34 *SD* of IMPACT scores) and student achievement (0.08 *SD*).
- When teachers identified by IMPACT as Effective or Highly Effective exited, teaching quality and student achievement fell, although the effects on student achievement were statistically insignificant.
- When teachers identified by IMPACT as low-performing (Ineffective or Minimally Effective) exited, teaching quality improved by 1.3 *SD*, and student achievement improved by 0.21 *SD* in math and 0.14 *SD* in reading, with nearly all gains accruing to students in high-poverty schools.

This pair of articles suggests that during its first 3 years, IMPACT induced compositional change that meaningfully improved academic outcomes for many of DCPS's poorest students. The latter article also indicates that IMPACT

evaluation ratings are aligned with student achievement outcomes.[3] However, while these results provide promising evidence of teacher evaluation's potential to improve teaching quality, there are several reasons why these effects may not persist once the system matures.

Changing Environment for Teacher Evaluation

The context for teacher evaluation nationally and in Washington, D.C., changed significantly since the early years of IMPACT. A growing public narrative paints teacher evaluation reform as a costly failure (Bill & Melinda Gates Foundation, 2018; Iasevoli, 2018; Strauss, 2015) and a waste of resources (Dynarski, 2016; National Council on Teacher Quality, 2017). In part, that assessment is informed by evidence outside of DCPS that evaluation has not meaningfully differentiated teacher effectiveness and few teachers are provided the information, incentives, or resources to improve or exit teaching (Kraft & Gilmour, 2017). For example, a recent study (Stecher et al., 2018) of three school districts and four charter management organizations found that teacher evaluation did not improve student achievement but also suffered from "incomplete implementation." There is also concern that high-stakes evaluation might dissuade entry into the profession, particularly for hard-to-staff schools (Kraft et al., 2019).

In the midst of a changing climate around teacher evaluation, DCPS made significant changes to IMPACT's design. The district (1) added, eliminated, and reweighted teaching-quality measures (in 2012–2013, 2014–2015, and 2016–2017) and (2) altered rating effectiveness bands (in 2012–2013). The range of IMPACT scores previously deemed Effective (250–349, out of a score range of 100–400) was divided in half; the upper half (300–349) remained Effective, but the lower half (250–299) was now labeled Developing. Teachers now identified as Developing (unlike their Effective peers) were subject to dismissal if they did not improve in 2 years. Additionally, DCPS (3) altered bonus and base pay incentives (2012–2013) and (4) introduced a performance-based career ladder (2012–2013). In 2014–2015, DCPS replaced its student achievement exam, DC CAS (Comprehensive Assessment System), with the PARCC (Partnership for Assessment of Readiness in College and Careers) exam. In 2016–2017, in response to low levels of achievement on PARCC, DCPS implemented LEAP (LEarning together to Advance our Practice), an intensive professional development program loosely coupled with IMPACT, which consists of 90-minute weekly small-group seminars and biweekly individual coaching. In addition, after 6 years of leadership by Kaya Henderson, DCPS has since 2016 had two interim chancellors and two new permanent chancellors, with additional turnover of deputy chancellors. Each of these changes may have led to significant disruptions and altered IMPACT's effectiveness.

Despite the changed context for teacher evaluation nationally and at DCPS, IMPACT's incentives continue to induce low-performing teachers to exit at significantly higher rates than otherwise-similar teachers. Employing data in the years since the Dee and Wyckoff (2015) RD analysis (2012–2013 through 2015–2016), Dee et al. (2019) find that Minimally Effective teachers exit at a rate that is 40% greater than otherwise-similar Developing teachers, and Developing teachers exit at a rate that is 40% greater than otherwise-similar Effective teachers. For Minimally Effective teachers who are retained, performance increases on average by 27% of an *SD* relative to otherwise-similar Developing teachers—quite similar effects to those from the first 3 years of IMPACT. The incentives confronting teachers near low-performing IMPACT ratings thresholds continue to induce teachers to alter their behavior relative to otherwise-similar peers facing substantially different incentives. However, the RD results represent only a small share of DCPS teachers. The proportion of teachers who are less than Effective has declined over time, and RD estimates by design are specific to a narrow bandwidth of scores. These effects say little about IMPACT's effects on teachers whose ratings are more distal from these thresholds and, *more specifically, do not imply that the increased turnover resulting from the exit of low-performing teachers improves teaching skills or student achievement.* That outcome depends on the differential effectiveness of leaving versus entering teachers.

Exiting teachers in recent years are likely to be more effective than exiting teachers from IMPACT's early years for several reasons. First, most of the lowest performing DCPS teachers at the inception of IMPACT have now exited—voluntarily or involuntarily. Second, many teachers whose scores would previously have designated them as Effective now fall in the Developing range; as such, they face a credible dismissal threat, which increases voluntary attrition (Dee et al., 2019). Third, changes in 2012–2013 to IMPACT's financial incentives for Highly Effective teachers concentrated incentives in high-poverty schools, which may increase turnover for the district's most effective teachers, who had been disproportionately situated in low-poverty schools. Fourth, teachers might find high-stakes evaluation stressful, which could increase the probability of turnover across the board; a concerning consequence would be if it induced exits among the most effective teachers who have disproportionately more attractive career alternatives (e.g., Feng & Sass, 2017). Finally, over the period of analysis in this article, the share of DCPS teachers who are Effective or Highly Effective has increased from 74% (2012–2013) to 82% (2017–2018). DCPS has likewise recently invested heavily in LEAP, a professional development program that is intended to improve the effectiveness of all teachers. Each of these mechanisms likely contributed to an increase in the measured performance of exiting teachers.

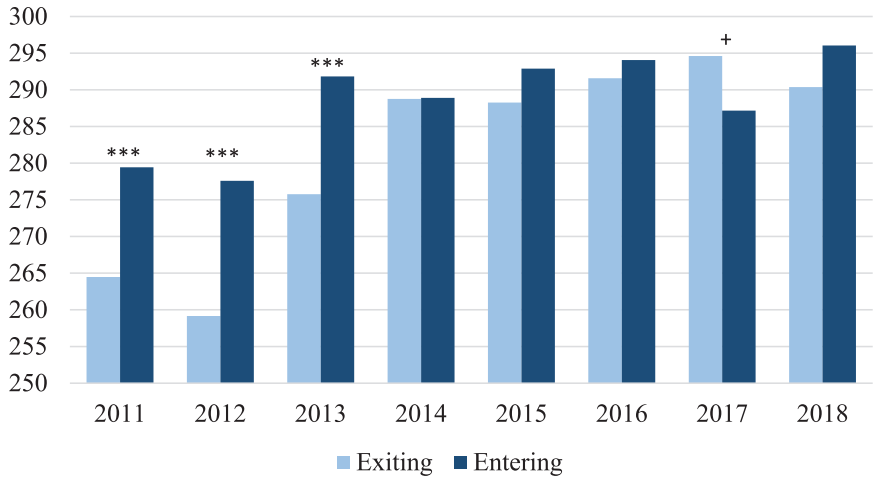

FIGURE 2 *Average IMPACT scores, all entering/exiting general education teachers, by year of replacement.*
Note. Teachers are assessed on IMPACT with scores from 100 to 400. Entering scores are averages for teachers new to DCPS at the end of their first year of teaching; exiting scores are those for teachers in the year before not returning to DCPS to teach. DCPS = District of Columbia Public Schools.
$^{***}p < .001$. $^{**}p < .01$. $^{*}p < .05$. $^{+}p < .10$. Represent statistical differences between entering/exiting teachers in each year.

External factors may likewise influence the quality of entering teachers. DCPS draws applicants from the larger market for teachers in the DC metropolitan area, and DC has a robust charter school presence. These charter schools could serve as a source of more effective, experienced teachers from which DCPS can draw. By design, teaching in DCPS brings the potential for atypically high salaries, which anecdotal reports indicate put pressure on charter schools to retain teachers (Brown, 2013). The share of teachers hired by DCPS with at least 3 years of experience has increased from 37% at IMPACT's inception in 2009–2010 to 62% in 2017–2018. These factors contribute to a strong applicant pool, although DCPS has not hired the teachers who are predicted to be the most effective (Jacob et al., 2018). Concerns over teacher accountability and teacher evaluation (Kraft et al., 2019), as well as a tight labor market (Taylor, 2019) may reduce the pool of applicants. The net effect of these competing mechanisms on the quality of entering teachers is conceptually unclear.

Empirically, descriptive evidence indicates that the effectiveness of both exiting and entering teachers increased in DCPS in recent years (Figure 2). The average IMPACT scores of exiting teachers in 2010–2011 and 2011–2012 was 262. Between 2012–2013 and 2017–2018, the average was 288, an increase of nearly 0.6 *SD*. Entering teachers' average IMPACT scores likewise improved but not by as much. The average differential between entering and exiting teachers in 2010–2011 and 2011–2012 was 17 IMPACT points; for 2012–2013 through 2017–2018 the average yearly difference was 4 points. These results call into question whether the effects of DCPS teacher turnover on student achievement found in Adnot et al. (2017) have been sustained. We examine this in detail below.

Method and Data

Our analysis comprises two parts. We first examine the causal effect of teacher turnover on teaching skills and student achievement, overall and differentiated by teacher effectiveness. Second, we descriptively examine the relationship between teacher evaluation and teacher turnover in DCPS.

The Effect of Teacher Turnover

We examine the effects of teacher exits from DCPS on teaching skills and student achievement by employing a panel-based design similar to prior research (Adnot et al., 2017; Chetty et al., 2014; Ronfeldt et al., 2013).[4] This design compares the effect of levels of teacher turnover in school grade cells on teaching quality and student achievement in year t with these outcomes in the same school grade cells in $t + 1$. Our analysis draws on DCPS administrative data from 2012–2013 through 2017–2018. To estimate the effect of teacher turnover on student achievement, we restrict the data to teachers who can be linked to student test scores (i.e., Grades 4–8 math and reading), and then collapse the data to the school grade level.[5] Changes in teaching effectiveness reflect the average differences between exiting and entering teachers, the disruption that such turnover creates among school grade colleagues, and the proportion of teachers in a school grade cell who exit. Changes in student achievement depend on similar differences, as well as the effect of differences in teaching skills on student achievement.

We aggregate what is intrinsically a teacher-level analysis to the school grade level to mitigate two potential problems. First, turnover effects likely reach beyond an

individual classroom to other classrooms in the same grade (Ronfeldt et al., 2013); we allow our turnover effect estimates to capture disruption effects and changes in peer effects, in addition to the compositional effects of school grade turnover.[6] Second, aggregation to the school grade level mitigates potential internal validity threats, such as when more motivated parents of children in grades with turnover attempt to seek returning teachers, leaving new teachers with lower performing students. We then estimate two reduced-form equations, one for each outcome of interest—teaching quality (Equation 1a) and student achievement (Equation 1b). We estimate effects separately for reading and math. Changes in teaching quality ($\Delta\overline{TQ}_{sgt}$) are measured by changes in IMPACT scores and are a function of: 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, E_{sgt-1}; a year fixed effect, ω_t; and a random error term, ε_{sgt}.[7] IMPACT scores are a weighted average of multiple measures, including ratings of teachers' core professionalism, classroom observation ratings, and their value added to student achievement.[8] Changes in student achievement are measured by changes in average residualized[9] student test scores, $\Delta\bar{A}^*_{sgt}$, and are a function of: changes in the attributes of grade-level peers, $\Delta\bar{X}_{sgt}$, 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, E_{sgt-1}, a year fixed effect, ω_t, and a random error term, ε^*_{sgt}.

$$\Delta\overline{TQ}_{sgt} = \gamma_1 E_{sgt-1} + \omega_t + \Delta\varepsilon_{sgt}. \tag{1a}$$

$$\Delta\bar{A}^*_{sgt} = \Delta\bar{X}_{sgt}\beta_2 + \gamma_1 E_{sgt-1} + \omega_t + \Delta\varepsilon^*_{sgt}. \tag{1b}$$

Estimates from these models identify the effects of turnover through a difference-in-differences approach—that is, by controlling for time-invariant traits specific to school grade cells, time-varying characteristics across schools and grades, and student-level characteristics including prior achievement. For example, the change in student performance in a school grade cell before and after teacher turnover captures the turnover effect and the effect of other time-invariant influences. A second difference between school grade cells with and without turnover isolates the effect of those other time-varying factors. The difference of these two isolates the effect of turnover.

Nonetheless, the internal validity of the teacher turnover estimates in Equation 1 rests on several assumptions. First, our approach as defined above assumes that DCPS does not manipulate transfers within DCPS such that it biases our estimates. An example that violates this assumption occurs when filling a vacancy created by turnover; a principal might systematically transfer teachers to the open position according to their effectiveness. We also assume in Equation 1 that these teacher transfers have no achievement implications for the "sending" school grade cell (e.g., due to disruption in the quality of peer teachers). To address this concern, we condition on the prevalence of within-school transfers, S_{sgt-1}, and transfers across schools in the district, D_{sgt-1}. Specifically, S_{sgt-1} is the share of school grade–year teachers who exited Grade g math (reading) at the end of $t-1$ but remained in school s, while D_{sgt-1} is the share of teachers who transferred out school s at the end of $t-1$ but remained teaching in the district. On average, 48% of replacement teachers come from outside the DCPS system, 14% transfer across DCPS schools, and 38% transfer across subjects or grades within DCPS schools. These controls allow us to condition on the effects that turnover may have on school grade cells that "send" teachers elsewhere within the school or district.

Next, we assume that students do not sort to or from schools in response to teacher turnover in a way that is correlated with student achievement. We also assume there are no unobserved school grade factors correlated with turnover and student achievement (e.g., changes in principal effectiveness, which could influence both teacher turnover and student achievement). To address these challenges to internal validity, we conduct several robustness checks. First, recall that our first differencing eliminates time-invariant school effects; we include school-by-year fixed effects to address the potential for school-level changes over time. Second, to explore student sorting, we estimate auxiliary regressions predicting student attributes with teacher turnover. If turnover predicts student attributes, it would suggest such sorting. In general, we find little evidence of this occurring (see Appendix Table A1). Finally, the theory of change underlying IMPACT is that improving teaching quality improves student achievement. Our estimates of the effects of turnover on teaching quality and turnover on student achievement allow us to explore this mechanism.[10] As will be seen in Tables 1 and 2, where turnover is estimated to positively or negatively affect achievement, we observe an effect of turnover on observed teaching quality that is of the same sign and of roughly proportionate magnitude. This increases our confidence that our estimates of the effects of teacher turnover on student achievement reflect the hypothesis that improved teacher quality, as measured by IMPACT, is a mechanism for improved student achievement, and not the other factors influenced by teacher turnover.

We extend our analysis to examine heterogeneous effects of turnover by measured teaching quality. In place of the overall turnover effect, E_{sgt-1}, we test specifications where E^L_{sgt-1} denotes the proportion of students in each such cell whose teacher exited the DCPS teaching workforce and was low performing (i.e., Developing, Minimally Effective, or Ineffective), and E^H_{sgt-1} denotes the proportion of students taught by a high-performing (Effective or Highly Effective)

TABLE 1

Effect of Turnover on IMPACT Scores and Student Achievement, 2012–2013 Through 2016–2017

	Math				Reading			
	(1) IMPACT score	(2) Student achievement	(3) IMPACT score	(4) Student achievement	(1) IMPACT score	(2) Student achievement	(3) IMPACT score	(4) Student achievement
All exits	5.87	0.050*			15.48**	0.028		
	(5.55)	(0.025)			(6.09)	(0.022)		
High-performer exits			−44.74***	−0.011			−31.02***	−0.031
			(6.70)	(0.033)			(6.01)	(0.023)
Low-performer exits			50.52***	0.105***			59.21***	0.085***
			(7.36)	(0.033)			(6.69)	(0.032)
Student controls		X		X		X		X
Observations	870	870	870	870	840	840	840	840

Note. 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 percentage of students in a school grade–year cell who are Black, Hispanic, other non-White race/ethnicity, limited-English–proficient, special education, or free or reduced-price lunch–eligible. Robust standard errors (in parentheses) are clustered at the school grade level. Pretreatment (i.e., exit) years span 2012–2013 through 2016-2017.
***$p < .001$. **$p < .01$. *$p < .05$. $^{\dagger}p < .10$.

TABLE 2

Effect of Turnover on IMPACT Scores and Student Achievement by School Poverty Status, 2012–2013 Through 2016–2017

	Math				Reading			
	(1) IMPACT score	(2) Student achievement	(3) IMPACT score	(4) Student achievement	(1) IMPACT score	(2) Student achievement	(3) IMPACT score	(4) Student achievement
All exits								
Low poverty	−7.82	0.076†			−12.72*	−0.036		
	(10.78)	(0.043)			(6.29)	(0.027)		
High poverty	7.39	0.047†			20.92***	−0.048		
	(5.88)	(0.027)			(6.78)	(0.132)		
High performers								
Low poverty			−29.56*	0.064			−21.23***	−0.017
			(13.38)	(0.048)			(6.71)	(0.036)
High poverty			−48.37***	−0.029			−35.23***	−0.036
			(7.31)	(0.037)			(7.52)	(0.027)
Low performers								
Low poverty			N/A	N/A			N/A	N/A
High poverty			49.51***	0.104***			60.17***	0.093***
			(7.56)	(0.034)			(6.92)	(0.033)
Student controls		X		X		X		X
Observations	870	870	870	870	840	840	840	840

Note. 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 percentage of students in a school grade–year cell who are Black, Hispanic, other non-White race/ethnicity, limited-English–proficient, special education, or free or reduced-price lunch–eligible. Robust standard errors (in parentheses) are clustered at the school grade level. Pretreatment (i.e., exit) years span 2012–2013 through 2016–2017. We do not include estimates for low-performer exits in low-poverty schools as these are found in only five school grade year observations during this period.
***$p < .001$. **$p < .01$. *$p < .05$. $^{\dagger}p < .10$.

teacher who left the district at the end of year $t-1$. The resulting specification takes the following form:

$$\Delta\overline{A}^{*}_{sgt} = \gamma_1 E^{L}_{sgt-1} + \gamma_2 E^{H}_{sgt-1} + \delta S_{sgt-1} + \theta D_{sgt-1} + \Delta\overline{X}_{sgt}\beta_2 + \omega_t + \varepsilon^{*}_{sgt}. \quad (2a)$$

$$\Delta\overline{TQ}_{sgt} = \gamma'_1 E^{L}_{sgt-1} + \gamma'_2 E^{H}_{sgt-1} + \delta' S_{sgt-1} + \theta' D_{sgt-1} + \omega'_t + \varepsilon'_{sgt}. \quad (2b)$$

For example, γ_1 in Equation 2a represents the effect of exits among low-performing teachers—those rated Developing, Minimally Effective, and Ineffective in year

$t-1$—on changes to student achievement in year t. We also repeat these analyses at the rating level to compare, for example, the effects of Highly Effective teachers' exit to that of Effective teachers.

Finally, we examine whether turnover effects vary by school poverty status by interacting each treatment variable with a school poverty indicator. We use DCPS's school poverty designation: any school with at least 60% of students eligible for free or reduced-price lunch is considered a high-poverty school. Close to 80% of the schools in our sample meet this definition.

Teacher Evaluation and Teacher Turnover

To explore the relationship between teacher evaluation and teacher turnover in DCPS we use administrative data and teacher responses to two surveys. First, rates of turnover are assessed using the administrative data described above. We define a teacher as having exited teaching in DCPS if they are a teacher of record earning an IMPACT score in t, and not a teacher of record in $t+1$ and $t+2$.[11] Teachers are linked to IMPACT and school data, facilitating analysis of teacher turnover by teaching effectiveness and student attributes using simple descriptive statistics.

We employ teachers' responses to two different surveys, linked to the administrative data, to assess teachers' views on teacher evaluation in DCPS and, for exiting teachers, the factors most important to their decision to exit. By linking these data sets, we can understand the reasoning underlying exit decisions for all DCPS general education teachers, as well as the same sample of teachers in tested grades and subjects who we analyze in Equations 1 and 2.[12] The first survey is the Declaration of Intent Not to Return (DINR), which all teachers who expect to leave DCPS are requested to complete by April 3.[13] Questions on the DINR have since 2012–2013 consistently asked teachers to rank their three primary reasons for leaving. We match DINR responses to 52% of all exits. The characteristics of DINR survey respondents often differ from the population of exiting teachers. Many of these differences are small, but two important differences should be noted. DINR respondents are higher performing and more likely to be White than exiting teachers as a whole (Appendix Table A2).

The second survey is Insight, which has been administered by TNTP to all DCPS teachers since 2015–2016, and asks a variety of questions ranging from teachers' experiences with professional development to their views on evaluation and their reasons for leaving or remaining in their schools. We use Insight responses to confirm the robustness of DINR responses to a different measure, as well as to assess DCPS teachers' views on evaluation. Insight survey responses were obtained from 95% of all teachers.

For both surveys, we stratify responses according to teachers' performance levels in order to understand how their responses vary according to teaching quality. Both surveys are administered in advance of teachers knowing their final IMPACT ratings for the year.

Analysis

About 20% of all DCPS teachers leave in a given year, and more than 50% exit over a 5-year period (Figure 1). These statistics raise many questions. Does this level of turnover negatively affect students? Which teachers leave DCPS, and why? Are they dissatisfied with policies or practices that could be readily changed? We explore these questions below.

Teacher Turnover, Teacher Quality, and Student Achievement

DCPS teacher turnover, on average, does not negatively affect teaching skills or student achievement. Student math achievement increases by an average of 5% of an *SD* in grades with turnover. The effect on reading is also positive, though estimates are statistically insignificant (see Table 1 and Figure 3). The overall effect, however, masks substantial heterogeneity with respect to the effectiveness of exiting teachers. The attrition of high-performing (Highly Effective or Effective) teachers reduces teaching quality, with suggestive evidence that student achievement also declines. When high-performing teachers leave, teaching skills in that grade decline on average by 31 (English language arts) and 45 (math) IMPACT points—equivalent to 0.75 and 0.96 *SD*s, respectively.[14] The effect of high-performing teachers' exits on student achievement is negative but never statistically significant. However, when we isolate the effect of turnover by the highest rated (Highly Effective) teachers, teaching skills decline by more than 1.5 *SD*, and student achievement in reading declines by 0.13 *SD* (Appendix Table A3). This represents more than 2 months of additional learning (Hill et al., 2008) and highlights the importance of retaining Highly Effective teachers. It likewise raises questions about how these teachers can be retained.

When low-performing (Ineffective, Minimally Effective, or Developing) teachers leave the classroom, on the other hand, we estimate substantial improvements to teaching quality and student achievement. Such exits improve teaching quality by more than 50 IMPACT points—equivalent to 1.1 and 1.4 *SD*s of IMPACT scores in math and reading, respectively. Student achievement improves by 0.11 (math) and 0.09 (reading) SDs (Table 1 and Figure 3). The effects of low-performing teachers' exits represent large gains to students—an increase in learning of about two months. IMPACT is not mechanically responsible for most of these exits, as 63% are voluntary. However, IMPACT's identification of low-performing teachers meaningfully increases their likelihood of voluntarily exiting, as indicated by Dee et al. (2019).

The losses associated with high-performing teachers' attrition, as well as the gains attributable to low-performing teachers' exit, primarily accrue to students in high-poverty schools, as shown in Table 2. The attrition of high-performing teachers from a high-poverty school results in a reduction of teaching quality of 1 *SD* in reading and math (48 and 35 IMPACT points, respectively). Effects on student achievement are negative but not statistically significant. Meanwhile, the exit of low-performing teachers from a high-poverty school increases teaching quality by well over 1 *SD* and student performance by 0.10 (math) and 0.09 (reading) *SD*s of student achievement.

Robustness of Results. We run a series of tests and alternative statistical specifications to assess the internal validity of our estimates. First, if students with different background characteristics were systematically assigned to turnover classrooms (e.g., if low-performing students were assigned to classrooms with teacher turnover), our results may not reflect the causal effect of turnover on teaching quality and student outcomes. To examine this, we replace our outcome variable with student attributes (Appendix Table A1). The results indicate that little, if any, such sorting is occurring.

A second potential source of endogeneity not directly controlled for in our primary specifications would occur if there were time-varying changes within the school outside of those captured by first differencing. One such example would be changes in school leadership, which might influence both teaching quality and student achievement, in addition to affecting teacher turnover. Appendix Table A4 presents results from specifications that control for school fixed effects, as well as school-by-year fixed effects. In both cases, the estimates are of similar magnitude, though they are in some cases underpowered relative to our primary specification.

Finally, other concurrent mechanisms at play within a school might influence teaching quality and student achievement outside of turnover within a given school grade–year cell. For example, turnover in other grades affecting cross-grade collaboration within a school could bias our estimates. We employ two placebo tests, following Adnot et al. (2017), to confirm that such mechanisms are not driving our results. In the first test, we replace the independent variable (turnover in year t) with turnover in $t+1$. These results (Appendix Table A5) indicate that turnover in $t+1$ has no effect on outcomes in t. In a second test, we add to our analysis a control for turnover in an adjacent (i.e., the next higher) grade to see if turnover in other grades is predictive of changes in student achievement or teaching quality in a given grade year cell, and whether turnover effects for that grade year cell differ when conditioned on turnover in other grades. The results of these tests, which we limit to those grade levels within a school that have higher adjacent grades, are consistent with the main results and indicate no effects from turnover in adjacent grades (see Appendix Table A6).

Summary. These results support three conclusions. First, there are large differences in the effects of losing teachers identified by IMPACT as high- versus low-performing. The estimated differential effect on teaching quality is more than 2 *SD*s, and more than about half a year's worth of student achievement in reading. Second, while no teacher evaluation system will perfectly differentiate true teacher effectiveness,[15] IMPACT makes useful distinctions. The estimated the differential effects on teaching quality (i.e., IMPACT scores) align well with estimated student achievement differences. Perhaps more telling, when high-performing teachers leave the classroom, students experience a modest negative achievement effect, while low-performing teachers' exits result in substantially improved student achievement. We would be concerned if IMPACT rating categories, which carry meaningful stakes, did not result in meaningful differences in student achievement.[16] Finally, the differential effects on student outcomes stemming from high- and low-performing teachers' exits suggest that the goal of improving teaching quality and student achievement is better served by retention strategies that target high-performing teachers rather than across-the-board retention efforts. We now turn to an analysis of the level and nature of teacher turnover in DCPS.

Teacher Effectiveness and Teacher Retention

Annual teacher attrition from DCPS averages just under 20%, with 5-year attrition averaging 57% from 2012–2013 through 2016–2017 (Figure 1). How should we contextualize this level of turnover? Overall, it is greater than a group of urban districts where 1- and 5-year teacher attrition averages 13% and 43%, respectively (Papay et al., 2017). However, given the differential effects of turnover on teaching quality and student achievement (Table 1, Figure 3) in DCPS, concern over teacher turnover should be informed by the effectiveness of exiting teachers.

The overall DCPS teacher turnover rate masks substantial variation across evaluation ratings (Figure 4). The least effective teachers are the most likely to exit in any given year. Between 2012–2013 and 2016–2017, attrition among Highly Effective teachers (11%) is a fifth as high as attrition among Minimally Effective teachers (55%) and less than half the rate for those rated Developing (26%). Among Highly Effective and Effective teachers—presumably the teachers DCPS is most interested in retaining—attrition is 13%. The differential turnover rates across levels of teacher effectiveness in DCPS are larger than have been documented in other, lower stakes settings (e.g., Feng & Sass, 2017; Goldhaber et al., 2010; Papay et al., 2017). However, given that the distribution of teaching quality in DCPS has

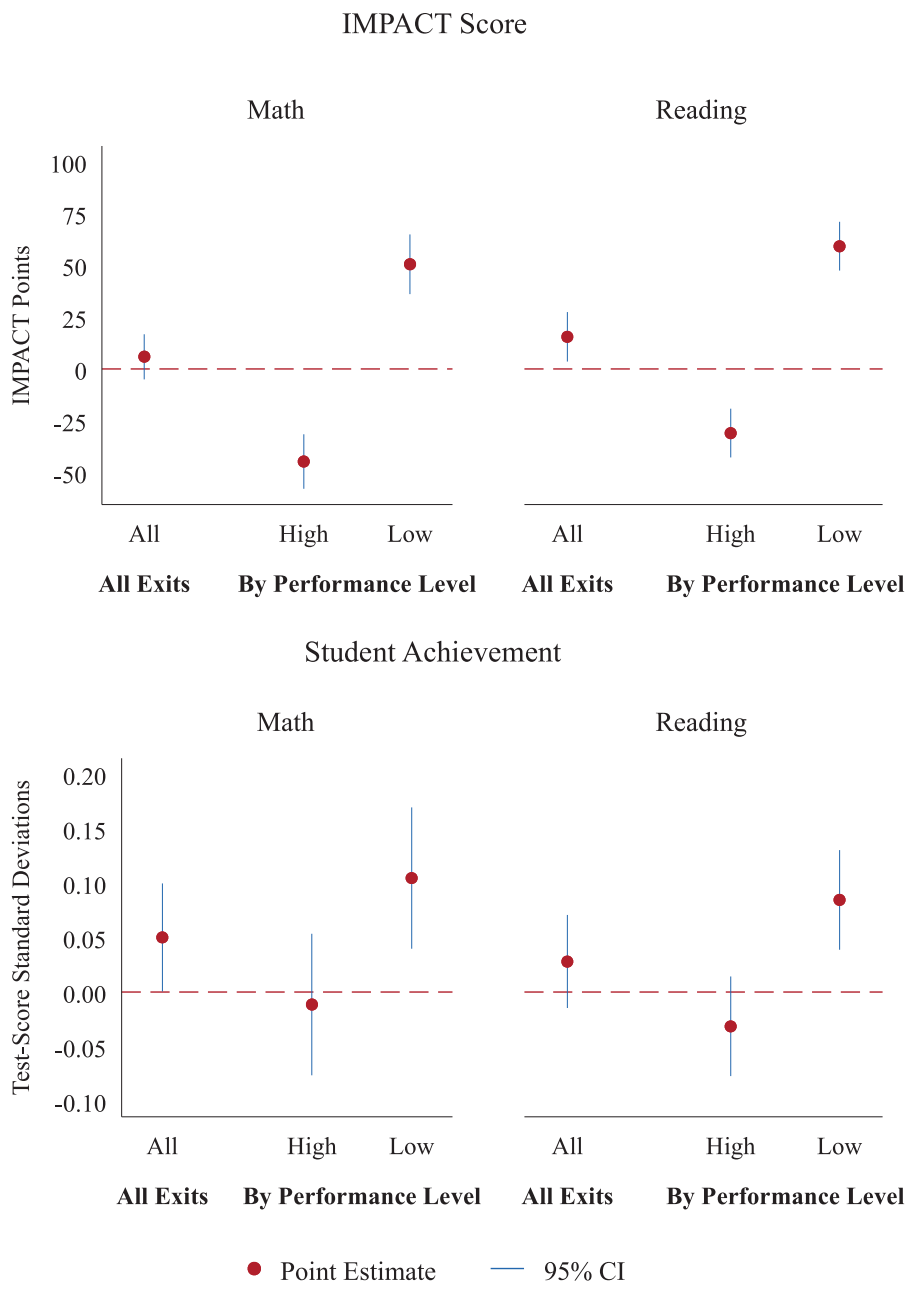

FIGURE 3 *Effects of teacher turnover on student achievement and teaching quality.*
Note. Models include year fixed effects and controls for teacher movement within and across schools. Student achievement estimates additionally control for the year-to-year, across-cohort change in the percentage of students in a school grade–year cell who are Black, Hispanic, other non-White race/ethnicity, limited-English–proficient, special education, or free or reduced-price lunch–eligible. CI = confidence interval.

improved over time (Dee et al., 2019), the relatively low rate of exit within high-performing teachers accounts for a meaningful share of overall exits. Slightly less than half (46%) of all DCPS attrition is among low-performing teachers during the 2012–2013 through 2016–2017 period. Policies that retain these teachers would reduce student achievement. However, 54% of DCPS turnover consists of high-performing teachers; 19% of DCPS turnover is from Highly Effective teachers. Although the turnover rate within Highly Effective teaches is relatively low (11%), losing these teachers comes at a cost to student achievement (Appendix Table A3), and policies should attempt to retain these teachers. In this regard, it is helpful to understand the reasons teachers cite as important to their exit decisions.

Teachers leave their positions for a variety of personal and professional reasons. Figure 5 summarizes the top factors (teachers can choose up to three) cited by DINR respondents between 2012–2013 and 2016–2017 in their decision

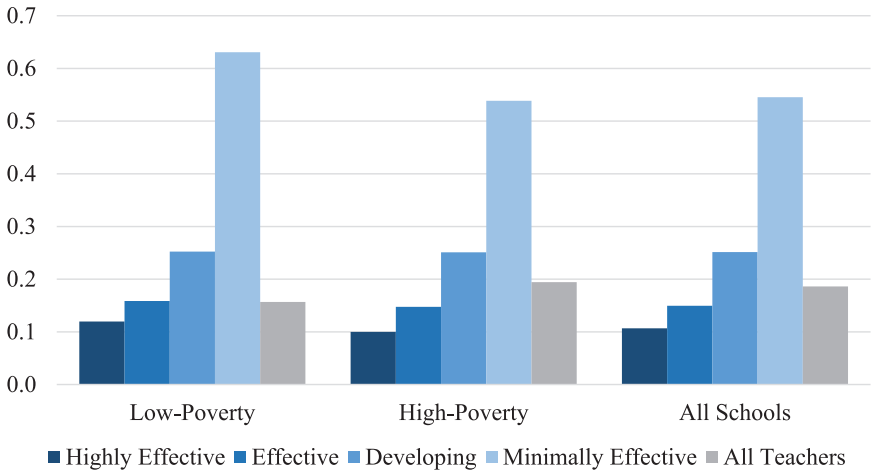

FIGURE 4 *Proportion of teachers exiting District of Columbia Public Schools by teacher performance level and school poverty, 2012–2013 through 2016–2017.*

Note. For all years except 2017, a teacher is retained if they taught and received an IMPACT rating in t and were a classroom teacher with an IMPACT rating in $t + 1$ or $t + 2$. For 2016–2017 teachers, retention includes those who were a classroom teacher with an IMPACT rating in $t + 1$ (2017–2018).

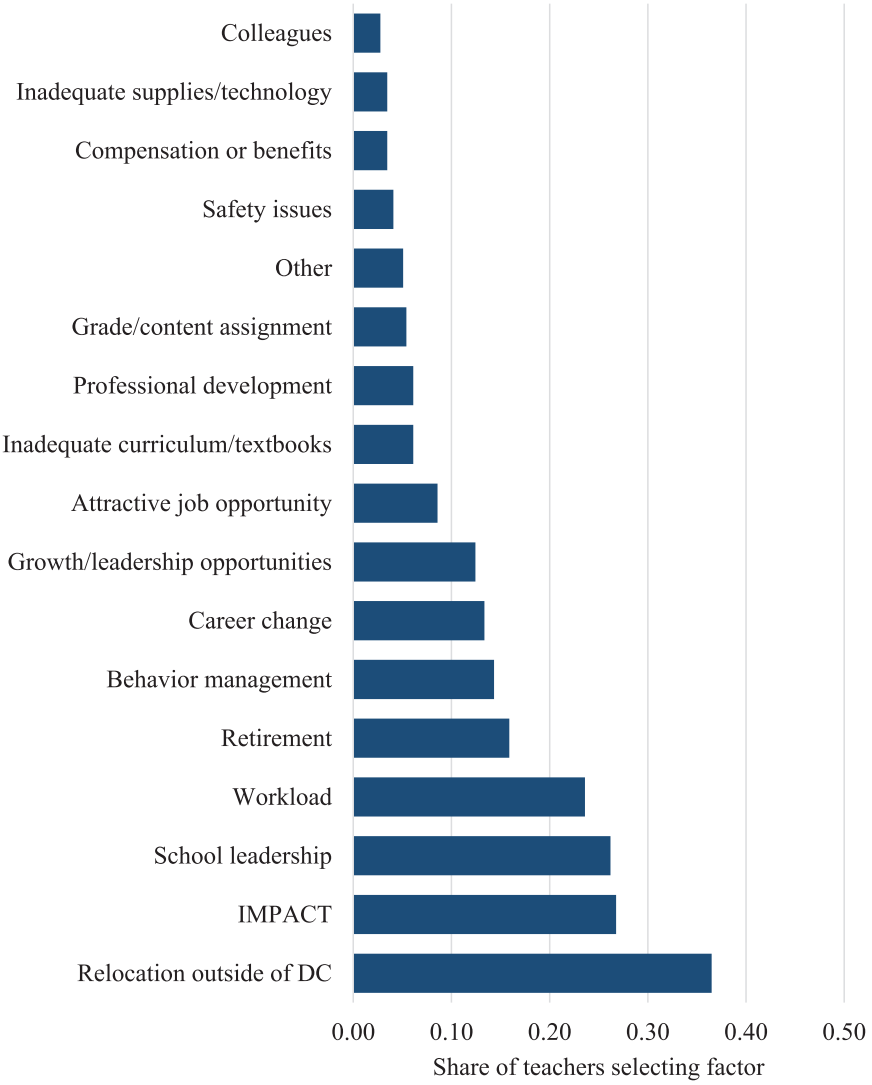

FIGURE 5 *(continued)*

FIGURE 5 *Factors identified as one of top three in decision to leave teaching in DCPS, from all exiting teachers' DINR survey responses, 2013-2017, all teachers.*
Note. Sample comprises responses to the DINR survey from teachers who gave notice to DCPS that they did not intend to continue teaching. Teachers were asked to select and rank up to 3 core factors in their decision to leave DCPS; the same teacher may therefore count toward multiple factors. The sample is restricted to exclude respondents who remained teaching in DCPS or whose retention status cannot yet be confirmed (i.e., respondents from the 2017–2018 academic year). $N = 1{,}737$. DCPS = District of Columbia Public Schools; DINR = Declaration of Intent Not to Return.

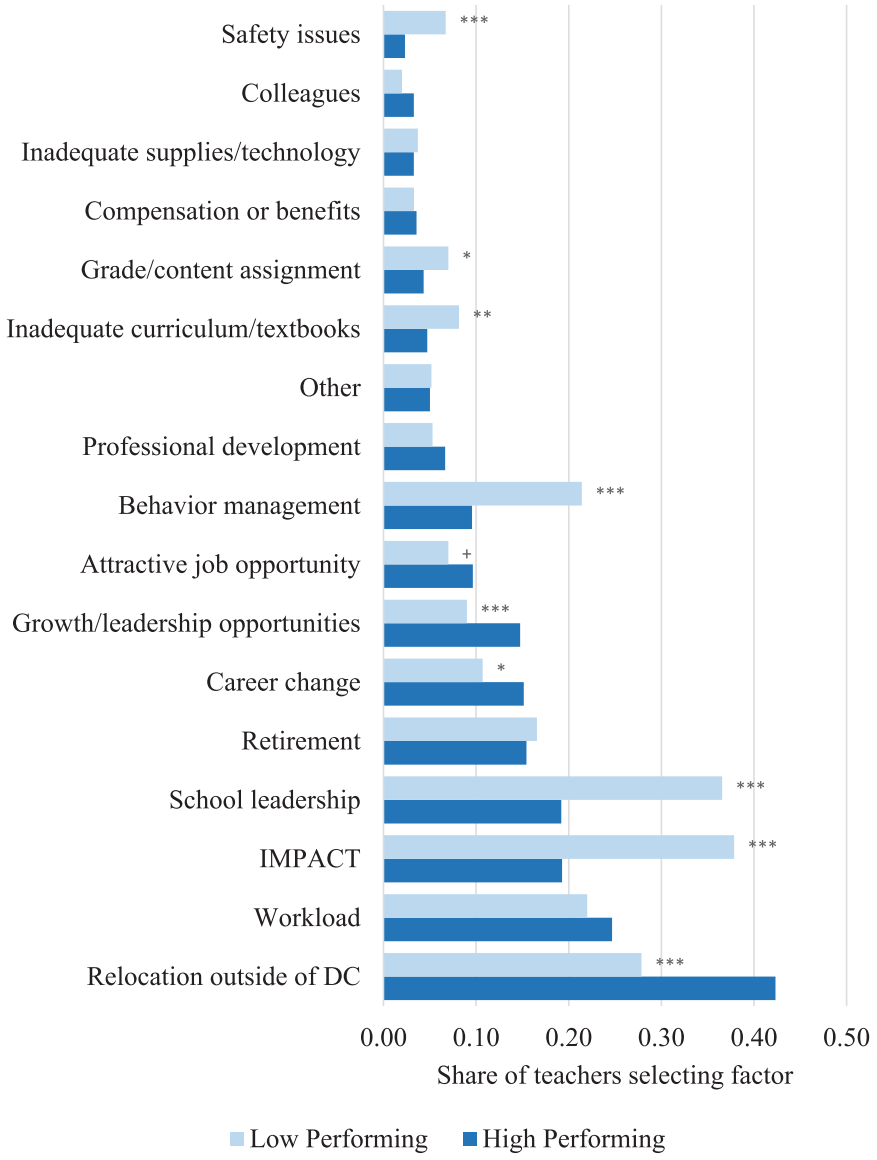

FIGURE 6 *Factors identified as one of top three in decision to leave teaching in DCPS, by teacher effectiveness, DINR survey responses, 2013-2017.*
Note. See note to Figure 5. High-Performing includes Effective and Highly Effective teachers. Low-Performing includes Developing and Minimally Effective teachers. DCPS = District of Columbia Public Schools; DINR = Declaration of Intent Not to Return.
$^{***}p < .001$. $^{**}p < .01$. $^{*}p < .05$. $^{+}p < .10$.

to exit the district.[17] The most commonly cited factor in teachers' decision to leave is relocation from the DC area (selected by 37%), followed by IMPACT (27%), school leadership (26%), and workload (24%). In contrast to other states and districts, teachers who leave DCPS rarely cite insufficient resources and inadequate pay; fewer than 5% of teachers describe compensation and benefits or the adequacy of school supplies as factors in their decision to leave.

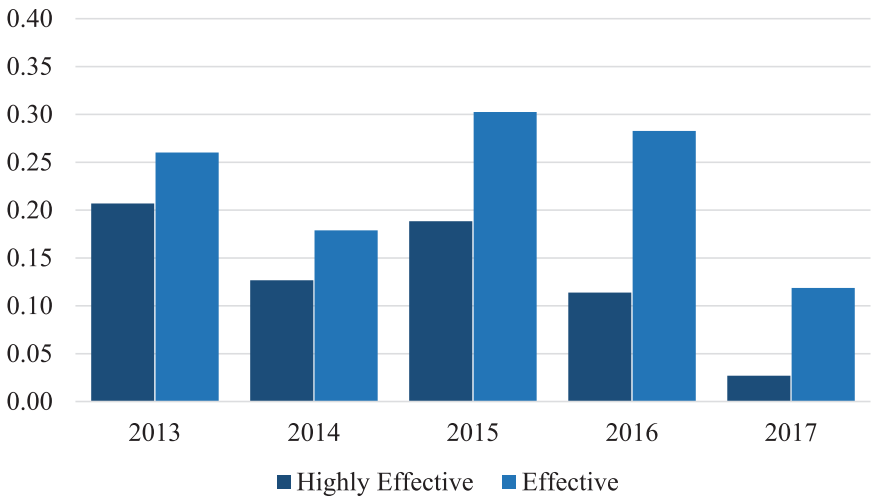

FIGURE 7 *Share of effective and highly effective teachers defining IMPACT as one of the top three factors in their decision to leave teaching in District of Columbia Public Schools, by year.*
Note. See note to Figure 3.

Teachers' reasons for leaving, however, differ by teacher performance (Figure 6). High-performing teachers are considerably more likely than low-performing teachers to say that they are leaving DCPS because they are relocating outside of DC (42% vs. 28%) or for growth and leadership opportunities (15% vs. 9%). Low-performing teachers, on the other hand, are roughly twice as likely as high-performing teachers to cite dissatisfaction with school leadership (37% vs. 19%) and concerns about behavior management (21% vs. 10%). Similarly, low-performing teachers are twice as likely as high-performing teachers (38% vs. 19%) to indicate that IMPACT was influential in their exit decisions. We explore the relationship between teacher turnover and teacher evaluation in greater detail below.

Although teacher evaluation has the potential to improve student outcomes, concerns have been raised that effective teachers find rigorous, high-stakes teacher evaluation so stressful that it affects their job satisfaction, driving exits from DCPS (Stein, 2019a, 2019b). This is a reasonable concern. Between 2012–2013 and 2016–2017 IMPACT 23% of exiting Effective teachers and 13% of exiting Highly Effective teachers cited IMPACT as one of their top three reasons for exiting (available from authors). The importance of IMPACT in teacher turnover has, however, diminished over time (Figure 7). IMPACT was named one of the top three factors by 21% of exiting Highly Effective teachers in 2013 but only 3% in 2017. Effective teachers identify IMPACT as a factor more frequently than Highly Effective teachers but also cite it less frequently over time. Even when high-performing (Highly Effective and Effective) teachers cited IMPACT as a factor in their decision to leave, it was rarely the primary factor. In 2017, for example, less than 2% of high-performing teachers identified IMPACT as their top-ranked reason for leaving DCPS (not shown).

More generally, teachers report favorably on the DCPS teacher evaluation system (Figure 8). A majority of all teachers, regardless of performance level, report somewhat to strong agreement that expectations for effective teaching are clearly defined at their schools (80%), that they know (90%) and agree with (69%) the performance criteria, that ratings are accurate reflections of their effectiveness (65%), and that evaluation helps them identify strengths and weaknesses (69%). There is, however, heterogeneity in these views. High-performing teachers demonstrate considerably higher agreement with the validity of these ratings than do low-performing teachers. Highly Effective teachers, whose buy-in might be critical for their retention, have high rates of agreement with the usefulness of the evaluation process for identifying their strengths and weaknesses (76%), largely agree that evaluation ratings are accurate reflections of their teaching effectiveness (78%), and agree with the criteria used to evaluate their performance (78%). Highly Effective teachers, whose teaching is most critical to the success of students, generally support IMPACT.

Discussion

Teacher turnover is much discussed by policymakers, researchers, and the popular media. These discussions usually present retention as unambiguously positive. However, a good portion of teacher turnover in DCPS works to the advantage of students, resulting in small improvements overall to student achievement in math and teacher quality in reading. These net positive effects can be explained by differential rates of turnover across levels of teaching quality and the ability of DCPS

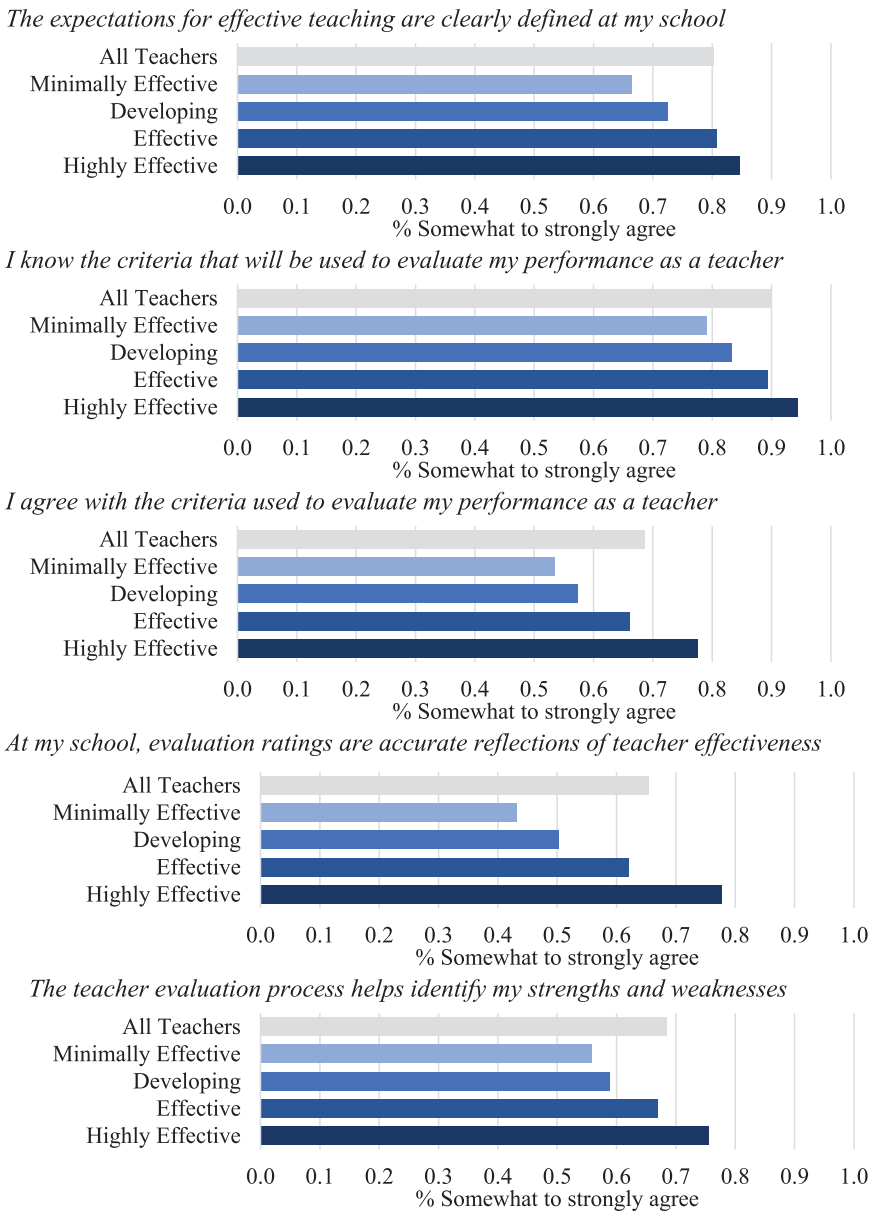

FIGURE 8 *Teachers views on teacher evaluation, by performance level.*
Note. Sample comprises responses to The New Teacher Project Insight survey from all teachers rated Minimally Effective through Highly Effective between 2015–2016 and 2017–2018. Each item is answered on a 6-point Likert-type scale ranging from *strongly disagree* to *strongly agree*. $N = 4{,}823$ unique teachers.

to hire from a relatively effective pool of replacements. Forty-six percent of teacher turnover from 2012–2013 through 2016–2017 was by teachers rated by IMPACT as Ineffective, Minimally Effective, or Developing. On average, when these teachers leave, they are replaced by teachers whose higher effectiveness leads to grade-level IMPACT scores that are almost an *SD* higher and additional student learning equivalent to about 2 months more in math or reading each year. As our results evidence, DCPS hires relatively effective replacement teachers. This is crucial and may not be the case in some other teacher labor markets.

However, our analysis also shows that Highly Effective teachers' exits—which account for 20% of all turnover in DCPS—can be costly to students. When such teachers exit, they are typically replaced by teachers whose IMPACT scores are about 1.5 *SD*s worse; in reading, this leads to a reduction in student learning of about 2 months. This places a premium on understanding why these teachers leave and developing policies and practices to moderate the losses. Although anecdotal concerns have been raised that IMPACT might drive teachers out of DCPS (see, e.g., Stein, 2019b), DCPS's best teachers hold

positive views of IMPACT and infrequently cite it as a reason for leaving; over a 5-year period, about 13% of exiting Highly Effective teachers cited IMPACT as among their top three reasons for leaving DCPS, and that number has declined substantially over time. Far fewer of these teachers rank IMPACT as the primary factor in their decision to leave (15% overall, 2% of exiting Highly Effective teachers). Additionally, of the factors within DCPS's grasp, Highly Effective teachers cite workload (25%), school leadership (18%), and opportunities for growth/leadership (15%) most frequently.

IMPACT is at an important juncture. The DC City Council has discussed legislation that could subject IMPACT to collective bargaining (Stein, 2019a), and DCPS Chancellor Lewis Ferebee announced recently that he is conducting a review of IMPACT to explore modifications that would improve teacher evaluation in DCPS (Stein, 2019b). This article, with Dee et al. (2019), shows that teaching and learning gains resulting from teacher evaluation can be sustained once these systems mature, even in the face of transitions in leadership, meaningful design modifications, implementation fatigue, competing priorities, and pressure from stakeholders. Nonetheless, about a quarter of Highly Effective teachers have concerns over the ability of IMPACT to reflect their effectiveness and identify their strengths and weaknesses, indicating room for improvement on measures of effectiveness. Although some aspects of IMPACT work well, it is likely that changes to IMPACT can at least partially address concerns of key stakeholders without jeopardizing these benefits.

What might the next phase of improvement look like? Improving performance measures to better align with policy goals is an ongoing process. Are there other or better measures? Evaluation tools are still being refined to improve their validity and reliability as measures of teaching quality and effectiveness. For example, are there ways to improve the reliability of teacher observations? Understanding which elements of the evaluation process are crucial to improving teaching quality of requires more research; one important and unanswered question, for example, is whether high stakes are necessary.

Given the unquestioned importance of effective teaching for student outcomes, teacher evaluation should focus on how best to identify teachers' effectiveness, and when weaknesses are identified, how best to support teachers' improvement. In a small percentage of cases, it may be necessary to dismiss teachers whose performance is sufficiently costly to students. Finally, this article demonstrates that evaluation might be employed more deliberately to retain high-performing teachers. This could include providing high-performing teachers with increased professional opportunities that leverage their skills—such as mentorship of other teachers (Papay et al., 2020)—and with direct communication about their value to the school, what can be done to retain them, and how to make better use of their skills.

Appendix A

TABLE A1
Effect of Teacher Turnover on Changes in Student Demographics

	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)
						Math						
All exits	0.013		0.002		−0.001		0.003		−0.006		0.007†	
	(0.009)		(0.007)		(0.009)		(0.008)		(0.008)		(0.004)	
High-performer exits		0.007		0.007		−0.011		−0.003		−0.003		0.008
		(0.013)		(0.009)		(0.009)		(0.010)		(0.009)		(0.006)
Low-performer exits		0.019†		−0.003		0.008		0.007		−0.009		0.006†
		(0.011)		(0.008)		(0.013)		(0.010)		(0.010)		(0.003)
Observations	870	870	870	870	870	870	870	870	870	870	870	870
						Reading						
All exits	0.020*		−0.005		−0.001		0.001		−0.012		0.003	
	(0.009)		(0.006)		(0.011)		(0.008)		(0.008)		(0.004)	
High-performer exits		0.010		−0.001		−0.016		−0.004		−0.010		0.001
		(0.012)		(0.010)		(0.014)		(0.010)		(0.011)		(0.005)
Low-performer exits		0.030**		−0.009		0.013		0.005		−0.014		0.006
		(0.012)		(0.007)		(0.012)		(0.010)		(0.008)		(0.005)
Observations	840	840	840	840	840	840	840	840	840	840	840	840

Note. Models include year fixed effects and controls for teacher movement within and across schools. Robust standard errors (in parentheses) are clustered at the school grade level. Pretreatment (i.e., exit) years span 2012–2013 through 2016–2017. FRPL = free or reduced-price lunch; SpEd = special education; LEP = limited-English–proficient.
***$p < .001$. **$p < .01$. *$p < .05$. †$p < .10$.

TABLE A2
Sample Characteristics, 2012–2013 Through 2016–2017

	All DCPS teachers			Exiting teachers			
		By school poverty			By school poverty		
	All	High	Low	All	High	Low	DINR respondents
n	18,397	14,405	3,992	3,374	2,767	607	1,737
IMPACT score	324	318	344	294	287	327	304***
High-performing	0.76	0.73	0.90	0.54	0.48	0.77	0.60***
Group 1	0.15	0.15	0.17	0.18	0.17	0.21	0.19
Gender							
Female	0.75	0.75	0.76	0.70	0.69	0.73	0.73***
Missing	0.01	0.01	0.01	0.06	0.06	0.05	0.06
Race/ethnicity							
Black	0.50	0.56	0.27	0.44	0.48	0.24	0.33***
Hispanic	0.05	0.05	0.05	0.05	0.05	0.06	0.06*
White	0.32	0.26	0.57	0.29	0.24	0.52	0.38***
Other	0.04	0.04	0.05	0.04	0.03	0.05	0.04*
Missing	0.08	0.09	0.06	0.18	0.19	0.13	0.19
Education							
Graduate degree	0.71	0.70	0.75	0.62	0.61	0.67	0.61
Missing	0.02	0.02	0.01	0.09	0.10	0.07	0.06***
Years of experience							
0	0.07	0.07	0.05	0.06	0.07	0.04	0.06
1-3	0.21	0.21	0.20	0.27	0.27	0.27	0.34***
4-8	0.25	0.24	0.29	0.25	0.23	0.35	0.26
≥9	0.47	0.48	0.45	0.39	0.40	0.34	0.33***
Missing	0.01	0.01	0.00	0.03	0.03	0.00	0.02**
School type							
Education campus	0.20	0.23	0.09	0.21	0.23	0.09	0.20
Elementary school	0.48	0.45	0.58	0.43	0.40	0.54	0.45**
Middle school	0.10	0.10	0.11	0.13	0.12	0.15	0.15***
High school	0.18	0.18	0.19	0.20	0.20	0.18	0.18*
Other	0.03	0.03	0.03	0.04	0.04	0.04	0.01***

Note. Group 1 teachers are those who teach tested grades and subjects. Schools in the "other" category include special programs, STAY schools, which serve students 16 or older, and special-education schools. High-performing teachers are those rated Effective or Highly Effective. DCPS = District of Columbia Public Schools; DINR = Declaration of Intent Not to Return.
***$p < .001$. **$p < .01$. *$p < .05$. $^{\dagger}p < .10$. Difference between DINR respondents and all DCPS exiters.

TABLE A3
Effect of Teacher Turnover on IMPACT Scores and Student Achievement, by Rating

	Math		Reading	
	(1) IMPACT	(2) Student	(1) IMPACT	(2) Student
Highly Effective exits	−77.33*** (10.98)	0.002 (0.050)	−67.96*** (10.52)	−0.133*** (0.042)
Effective exits	−34.11*** (7.40)	−0.037 (0.038)	−25.98*** (6.69)	0.001 (0.028)
Student controls		X		X
Observations	870	870	840	840

Note. 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 percentage of students in a school grade–year cell who are Black, Hispanic, other non-White race/ethnicity, limited-English–proficient, special education, or free or reduced-price lunch–eligible. Robust standard errors (in parentheses) are clustered at the school grade level. Pretreatment (i.e., exit) years span 2012–2013 through 2016–2017. Models also include controls for low-performing teacher exits.
***$p < .001$. **$p < .01$. *$p < .05$. $^{\dagger}p < .10$.

TABLE A4

Effect of Teacher Turnover on IMPACT Scores and Student Achievement, Conditional on School Fixed Effects

	IMPACT score				Student achievement			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
				Math				
All exits	6.87	9.66**			0.059†	0.055**		
	(6.60)	(8.07)			(0.03)	(0.034)		
High-performer exits			−44.45***	−39.95***			−0.005	−0.019
			(7.90)	(10.03)			(0.04)	(0.037)
Low-performer exits			55.67***	55.69***			0.124***	0.129***
			(8.19)	(9.74)			(0.041)	(0.048)
Student controls					X	X	X	X
School fixed effect	X		X		X		X	
School-by-year fixed effect		X		X		X		X
Observations	870	870	870	870	870	870	870	870
				Reading				
All exits	17.85***	13.57**			0.034	0.051**		
	(6.87)	(7.58)			(0.026)	(0.031)		
High-performer exits			−30.52***	−26.82***			−0.027	0.026
			(6.87)	(8.00)			(0.027)	(0.034)
Low-performer exits			65.24***	52.74***			0.095**	0.076
			(7.44)	(7.21)			(0.038)	(0.050)
Student controls					X	X	X	X
School fixed effect	X		X		X		X	
School-by-year fixed effect		X		X		X		X
Observations	840	840	840	840	840	840	840	840

Note. 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 percentage of students in a school grade–year cell who are Black, Hispanic, other non-White race/ethnicity, limited-English–proficient, special education, or free or reduced-price lunch–eligible. Robust standard errors (in parentheses) are clustered at the school grade level. Pretreatment (i.e., exit) years span 2012–2013 through 2016–2017.
***$p < .001$. **$p < .01$. *$p < .05$. †$p < .10$.

TABLE A5

Effect of Next-Year Teacher Turnover on IMPACT Scores and Student Achievement

	Math				Reading			
	(1) IMPACT score	(2) Student achievement	(3) IMPACT score	(4) Student achievement	(1) IMPACT score	(2) Student achievement	(3) IMPACT score	(4) Student achievement
All exits	3.84	−0.050			0.04	0.011		
	(6.53)	(0.032)			(6.18)	(0.029)		
High-performer exits			−0.37	−0.081*			−2.76	0.063
			(7.51)	(0.037)			(7.56)	(0.041)
Low-performer exits			9.62	−0.007			2.83	−0.040
			(9.59)	(0.046)			(8.29)	(0.030)
Student controls		X		X		X		X
Observations	696	696	696	696	672	672	672	672

Note. Models include year fixed effects and controls for next-year teacher movement within and across schools. Student controls account for the year-to-year, across-cohort change in the percentage of students in a school grade–year cell who are Black, Hispanic, other non-White race/ethnicity, limited-English–proficient, special education, or free or reduced-price lunch–eligible. Robust standard errors (in parentheses) are clustered at the school grade level. Pretreatment (i.e., exit) years span 2012–2013 through 2015–2016, as next-year turnover is not available for 2016–2017.
***$p < .001$. **$p < .01$. *$p < .05$. †$p < .10$.

TABLE A6
Effect of Next-Grade Teacher Turnover on IMPACT Scores and Student Achievement

	Math				Reading			
	(1) IMPACT score	(2) Student achievement	(3) IMPACT score	(4) Student achievement	(1) IMPACT score	(2) Student achievement	(3) IMPACT score	(4) Student achievement
All exits, current grade	7.36	0.067**	9.13	0.072**	17.50*	0.012**	16.92*	0.013**
	(7.36)	(0.033)	(7.57)	(0.032)	(7.64)	(0.026)	(7.78)	(0.026)
All exits, next grade			−8.78	−0.027			2.79	−0.001
			(5.87)	(0.024)			(7.22)	(0.027)
Student controls		X		X		X		X
Observations	551	551	551	551	500	500	500	500

Note. Models include year fixed effects and controls for teacher movement within and across schools, for the subsample of nonterminal school grade–year cells. Student controls account for the year-to-year, across-cohort change in the percentage of students in a school grade–year cell who are Black, Hispanic, other non-White race/ethnicity, limited-English–proficient, special education, or free or reduced-price lunch–eligible. Robust standard errors (in parentheses) are clustered at the school grade level. Pretreatment (i.e., exit) years span 2012–2013 through 2016–2017.
***$p < .001$. **$p < .01$. *$p < .05$. $^{\dagger}p < .10$.

Acknowledgments

We are grateful to the District of Columbia Public Schools for supplying the data employed in this research and to Chris Lewis, Astrid Atienza, and Sooyon Stiller for answering our questions. We are likewise grateful to Allison Deptula and Kimberlie Schifrin at TNTP for sharing the Insight survey data with us. We appreciate helpful comments from John Papay and three anonymous reviewers on an earlier draft.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: We received financial support from the Schusterman Family Foundation, the Overdeck Family Foundation, the Institute of Education Sciences Grants R305H140002 and R305B140026. The views expressed in the article and any errors are attributable to the authors.

ORCID iD

Jessalynn James https://orcid.org/0000-0001-6765-2642

Notes

1. For recent summaries, see Simon and Johnson (2015), Kraft et al. (2016), and Papay et al. (2017).

2. For more details on the context and development of IMPACT, see Toch (2018).

3. IMPACT's classroom observation scores are correlated with teachers' value-added at 0.30 (Meyer, 2016). We also estimate that lagged IMPACT scores are predictive of current-year effects on student achievement, with correlations of approximately 0.30 in both subjects. These associations are comparable to other studies of teacher evaluation measures (e.g., Bell et al., 2012; Gill et al., 2016; Kane et al., 2013; Milanowski, 2004; Whitehurst et al., 2014).

4. This discussion draws directly on that presented in Adnot et al. (2017).

5. We follow the same dataset construction procedures as Adnot et al. (2017), creating unique datasets for math and reading. For a detailed description of these procedures, please refer to the online appendix to Adnot et al. (2017), available at https://journals.sagepub.com/doi/suppl/10.3102/0162373716663646.

6. Note that we are unable to link teachers to individual classrooms.

7. Changes in teaching quality and the share of exiting teachers are weighted by the share of students taught. These weights are employed so that point estimates can be interpreted in terms of the magnitude of turnover effects for the average student. There is a debate whether to control for student attributes when examining measures of teacher quality (e.g., Whitehurst et al., 2014; Steinberg & Garrett, 2016; Campbell & Ronfeldt, 2018). IMPACT includes controls for student characteristics in its value-added scores, but not for teacher observations. Because this is DCPS's measure of teacher quality, we do not include these controls in our main estimates; however, estimates from models that include these (available upon request) are nearly identical to results from models omitting student characteristics.

8. Every school-based DCPS employee receives an IMPACT score, regardless of role or tenure. For a brief overview of these measures and IMPACT's design, see Dee et al. (2019). Detailed descriptions are also available on DCPS's website at https://dcps.dc.gov/page/impact-dcps-evaluation-and-feedback-system-school-based-personnel.

9. We standardize test scores within subject, grade, and year, and then residualize for student characteristics such as race, poverty status, and special education status. Through 2013–2014, the test used in DCPS was the DC CAS; in 2014–2015 DCPS transitioned to the PARCC exam.

10. Dividing i from Equation 2a by γ_t' from Equation 2b approximates the Wald estimator, which represents the change in student achievement due to changes in teacher quality that result from teacher turnover.

11. All DCPS teachers present long enough in the school year to receive at least one classroom observation score receive an IMPACT score. Our focus is on classroom teachers, so someone who is a teacher in t but remains in DCPS in another capacity in

$t+1$ is considered as having exited. Teachers absent from a teaching role for more than 2 years are defined as having exited.

12. We present survey results for the full sample of respondents, though the subset of teachers in tested grades and subjects who comprise the analytic sample for the preceding analysis rank factors similarly to general education teachers broadly.

13. Those who leave the district before the subsequent school year and have not completed the DINR by April 3 will be fined $1,000.

14. One standard deviation of IMPACT for the analytic samples is equal to 46.7 points (math) and 40.9 points (English language arts).

15. All measurement systems suffer some degree of measurement error. Even a multimeasure system like IMPACT may not accurately capture all dimensions of effective teaching. Ultimately, we want such systems to be as correlated with true effectiveness as possible. That said, even an imperfect system that yields gains for students over the long run will be preferable to no system.

16. In addition, we test this mechanism by adding changes in teaching quality to our regressions estimating effects on student achievement (not shown). Indeed, the inclusion of changes in teaching IMPACT scores explains much of the variation in turnover effects on student achievement—for both subjects and turnover overall, as well as for high- versus low-performing exits.

17. We use Insight survey responses to test whether our DINR findings are robust to a separate survey instrument, and we find that teachers cite the same exit factors at comparable rates on both surveys (not shown).

References

Adnot, M., Dee, T., Katz, V., & Wyckoff, J. (2017). Teacher turnover, teacher quality, and student achievement in DCPS. *Educational Evaluation and Policy Analysis*, *39*(1), 54–76. https://doi.org/10.3102/0162373716663646

Bill Gates & Melinda Gates Foundation. (2018). *2018 Annual letter*. https://www.gatesfoundation.org/Who-We-Are/Resources-and-Media/Annual-Letters-List

Bell, C., Gitomer, D., McCaffrey, D., Hamre, B., Pianta, R., & Qi, Y. (2012). An argument approach to observation protocol validity. *Educational Assessment*, *17*(2-3), 62–87. https://doi.org/10.1080/10627197.2012.715014

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. H. (2008). *Who leaves? Teacher attrition and student achievement* (NBER Working Paper No. 14022). National Bureau of Economic Research. https://doi.org/10.3386/w14022

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2013). Analyzing the determinants of the matching of public school teachers to jobs: Disentangling the preferences of teachers and employers. *Journal of Labor Economics*, *31*(1), 83–117. https://doi.org/10.1086/666725

Brown, E. (2013, August 19). D.C. traditional public school teacher pay is higher than charters. *The Washington Post*. https://www.washingtonpost.com/local/education/dc-traditional-public-school-teacher-pay-is-higher-than-charters/2013/08/19/9fb3edf8-0518-11e3-9259-e2aafe5a5f84_story.html

Cambell, S., & Ronfeldt, M. (2018). Observational evaluation of teachers: Measuring more than we bargained for? *American Educational Research Journal*, *55*(6), 1233–1267. https://doi.org/10.3102/0002831218776216

Carver-Thomas, D., & Darling-Hammond, L. (2019). The trouble with teacher turnover: How teacher attrition affects students and schools. *Education Policy Analysis Archives*, *27*(36), 1–27. https://doi.org/10.14507/epaa.27.3699

Chetty, R., Friedman, J. H., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633–2679. https://doi.org/10.1257/aer.104.9.2633

Clotfelter, C., Glennie, E., Ladd, H., & Vigdor, J. (2008). Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina. *Journal of Public Economics*, *92*(5–6), 1352–1370. https://doi.org/10.1016/j.jpubeco.2007.07.003

Cohen, J., Loeb, S., Miller, L. C., & Wyckoff, J. H. (2020). Policy implementation, principal agency, and strategic action: Improving teaching effectiveness in New York City middle schools. *Educational Evaluation and Policy Analysis*, *42*(1), 134–169. https://doi.org/10.3102/0162373719893338

Cowan, J., & Goldhaber, D. (2018). Do bonuses affect teacher staffing and student achievement in high poverty schools? Evidence from an incentive for national board certified teachers in Washington State. *Economics of Education Review*, *65*(1), 138–152. https://doi.org/10.1016/j.econedurev.2018.06.010

Cullen, J., Koedel, C., & Parsons, E. (2019) The compositional effect of rigorous teacher evaluation on workforce quality. *Education Finance and Policy*. Advance online publication. https://doi.org/10.1162/edfp_a_00292

Dee, T. S., James, J., & Wyckoff, J. (2019). Is effective teacher evaluation sustainable? Evidence from DCPS. *Education Finance and Policy*. Advance online publication. https://doi.org/10.1162/edfp_a_00303

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

Dynarski, M. (2016). *Teacher observations have been a waste of time and money*. Brookings Institution. https://www.brookings.edu/research/teacher-observations-have-been-a-waste-of-time-and-money/

Feng, L., & Sass, T. (2017). Teacher quality and teacher mobility. *Education Finance and Policy*, *12*(3), 396–418. https://doi.org/10.1162/EDFP_a_00214

Gill, B., Shoji, M., Coen, T., & Place, K. (2016). *The content, predictive power, and potential bias in five widely used teacher observation instruments* (REL 2017-191). Regional Educational Laboratory Mid-Atlantic, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Glazerman, S., Protik, A., Teh, B., Bruch, J., & Max, J. (2013). *Transfer incentives for high-performing teachers: Final results from a multisite experiment* (NCEE 2014-4003). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/pubs/20144003/pdf/20144003.pdf

Glazerman, S., & Seifullah, A. (2012). *An evaluation of the Chicago Teacher Advancement Program (Chicago TAP) after four years*. Mathematica Policy Research. https://www.mathematica.org/our-publications-and-findings/publications/

an-evaluation-of-the-chicago-teacher-advancement-program-chicago-tap-after-four-years

Goldhaber, D., Gross, B., & Player, D. (2010). 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. https://doi.org/10.1002/pam.20549

Grissom, J., & Bartanen, B. (2019). Strategic retention: Principal effectiveness and teacher turnover in multiple-measure teacher evaluation systems. *American Education Research Journal*, *56*(2), 514–555. https://doi.org/10.3102/0002831218797931

Hanushek, E., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). *The market for teacher quality* (NBER Working Paper No. 11154). National Bureau for Economic Research. https://doi.org/10.3386/w11154

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

Iasevoli, B. (2018, February 15). *Teacher-evaluation efforts haven't shown results, say Bill and Melinda Gates*. http://blogs.edweek.org/edweek/teacherbeat/2018/02/teacher_evaluation_efforts_haven%27t_shown_results_bill_melinda_gates.html?cmp=soc-edit-tw

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*, *38*(3), 499–534. https://doi.org/10.3102/00028312038003499

Jacob, B. A., Rockoff, J. E., Taylor, E. S., Lindy, B., & Rosen, R. (2018). Teacher applicant hiring and teacher performance: Evidence from DC public schools. *Journal of Public Economics*, *166*(1), 81–97. https://doi.org/10.1016/j.jpubeco.2018.08.011

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment*. Bill & Melinda Gates Foundation. http://k12education.gatesfoundation.org/download/?Num=2676&filename=MET_Validating_Using_Random_Assignment_Research_Paper.pdf

Kraft, M., Brunner, E., Dougherty, S., & Schwegman, D. (2019). *Teacher accountability reforms and the supply and quality of new teachers* (EdWorkingPaper No. 19-169). Annenberg Institute at Brown University. https://doi.org/10.26300/7bcw-5r61

Kraft, M. A., & Gilmour, A. (2017). Revisiting *The Widget Effect*: Teacher evaluation reform and the distribution of teacher effectiveness. *Educational Researcher*, *46*(5), 234–249. https://doi.org/10.3102/0013189X17718797

Kraft, M. A., Marinell, W. H., & Shen-Wei Yee, D. (2016). School organizational contexts, teacher turnover, and student achievement: Evidence from panel data. *American Educational Research Journal*, *53*(5), 1411–1449. https://doi.org/10.3102/0002831216667478

Levy, M. (2018). *Teacher and principal turnover in public schools in the District of Columbia*. District of Columbia State Board of Education. https://sboe.dc.gov/sites/default/files/dc/sites/sboe/publication/attachments/SBOE%20Teacher%20Turnover%20Report%20-%20FINAL.pdf

Loeb, S., Miller, L. C., & Wyckoff, J. (2015). Performance screens for school improvement: The case of teacher tenure reform in New York City. *Educational Researcher*, *44*(4), 199–212. https://doi.org/10.3102/0013189X15584773

Meyer, J. P. (2016). *Reliability of and validity evidence for Teaching Learning Framework scores for the District of Columbia public school system* [Unpublished manuscript]. Curry School of Education, University of Virginia.

Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education*, *79*(4), 33–53. https://doi.org/10.1207/s15327930pje7904_3

Murnane, R. J. (1984). Selection and survival in the teacher labor market. *Review of Economics and Statistics*, *66*(3), 513–518. https://doi.org/10.2307/1925013

National Commission on Teaching and America's Future. (2003). *No dream denied: A pledge to America's children* [Summary report].

National Council on Teacher Quality. (2017). *Running in place: How new teacher evaluations fail to live up to promises*. https://www.nctq.org/publications/Running-in-Place:-How-New-Teacher-Evaluations-Fail-to-Live-Up-to-Promises

National Research Council. (2015). *An evaluation of the public schools of the District of Columbia: Reform in a changing landscape*. National Academies Press. https://doi.org/10.17226/21743

Papay, J. P., Bacher-Hicks, A., Page, L., & Marinell, W. H. (2017). The challenge of teacher retention in urban schools: Evidence of variation from a cross-site analysis. *Educational Researcher*, *46*(8), 434–448. https://doi.org/10.3102/0013189X17735812

Papay, J. P., Taylor, E., Tyler, J., & Laski, M. (2020). Learning job skills from colleagues at work: Evidence from a field experiment using teacher performance data. *American Economic Journal: Economic Policy*, *12*(1), 359–388. https://doi.org/10.1257/pol.20170709

Rockoff, J. E., Staiger, D. O., Kane, T. J., & Taylor, E. S. (2012). Information and employee evaluation: Evidence from a randomized intervention in public schools. *American Economic Review*, *102*(7), 3184–3213. https://doi.org/10.1257/aer.102.7.3184

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal*, *50*(1), 4–36. https://doi.org/10.3102/0002831212463813

Sartain, L., & Steinberg, M. P. (2016). Teachers' labor market responses to performance evaluation reform: Experimental evidence from Chicago public schools. *Journal of Human Resources*, *51*(3), 615–655. https://doi.org/10.3368/jhr.51.3.0514-6390R1

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

Smith, R., & Handler, H. (1979). Research on retention of teachers. *Review of Research in Education*, *7*(1), 418–443. https://doi.org/10.3102/0091732X007001418

Springer, M. G., Swain, W. A., & Rodriguez, L. A. (2016). Effective teacher retention bonuses: Evidence from Tennessee. *Educational Evaluation and Policy Analysis*, *38*(2), 199–221. https://doi.org/10.3102/0162373715609687

Stecher, B., Holtzman, D., Garet, M., Hamilton, L., Engberg, J., Steiner, E., Robyn, A., Baird, M. D., Gutierrez, I. A., Peet, E. D., de los Reyes, I. B., Fronberg, K., Weinberger, G., Hunter, G. P., & Chambers, J. (2018). *Improving teaching effectiveness: Final report: The intensive partnerships for effective teaching*

through 2015-16 [Policy report]. RAND Corporation. https://doi.org/10.7249/RR2242

Stein, P. (2019a, June 30). *With union backing, D.C. Council introduces proposed overhaul of controversial teacher evaluation system*. The Washington Post. https://www.washingtonpost.com/local/education/with-union-backing-dc-council-introduces-proposed-overhaul-of-controversial-teacher-evaluation-system/2019/06/29/f3722a7a-992f-11e9-8d0a-5edd7e2025b1_story.html

Stein, P. (2019b, October 21). Chancellor pledges to review D.C.'s controversial teacher evaluation system. *The Washington Post*. https://www.washingtonpost.com/local/education/chancellor-vows-to-review-the-districts-controversial-teacher-evaluation-system/2019/10/20/6c00405c-f0de-11e9-8693-f487e46784aa_story.html

Steinberg, M., & Garrett, R. (2016). Classroom composition and measured teacher performance: What do teacher observation scores really measure? *Educational Evaluation and Policy Analysis*, *38*(2), 293–317. https://doi.org/10.3102/0162373715616249

Strauss, V. (2015, January 1). Teacher evaluation: Going from bad to worse? *The Washington Post*.

Taylor, Y. S. (2019, June 19). *Growing labor demand in D.C. is driving up wages*. DC Policy Center. https://www.dcpolicycenter.org/publications/employment-trends-washington-region/

The New Teacher Project. (2012). *The irreplaceables: Understanding the real retention crisis in America's urban schools*. https://tntp.org/assets/documents/TNTP_Irreplaceables_2012.pdf

Toch, T. (2018). *A policymaker's playbook: Transforming public school teaching in the nation's capital*. Future Ed, Georgetown University. https://www.future-ed.org/wp-content/uploads/2018/06/APOLICYMAKERSPLAYBOOK.pdf

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The Widget Effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. The New Teacher Project. https://tntp.org/assets/documents/TheWidgetEffect_2nd_ed.pdf

Whitehurst, G. J., Chingos, M. M., & Lindquist, K. M. (2014). *Evaluating teachers with classroom observations: Lessons learned in four districts*. Brown Center on Education Policy at Brookings Institute. https://www.brookings.edu/wp-content/uploads/2016/06/Evaluating-Teachers-with-Classroom-Observations.pdf

Authors

JESSALYNN JAMES is a postdoctoral research associate at the Annenberg Institute at Brown University. She studies policies that influence teacher quality, specifically through teacher recruitment, retention, and development.

JAMES H. WYCKOFF is Curry Memorial Professor of Education and Public Policy at the University of Virginia. His research focuses on teacher labor markets, teacher quality, and policies to improve teaching in low-performing schools.