I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17576

On-the-Job Learning: How Peers and Experience Drive Productivity Among Teachers

Romaine A. Campbell
Seth Gershenson
Constance A. Lindsay
Nicholas W. Papageorge
Jessica H. Rendon

DECEMBER 2024

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17576

On-the-Job Learning: How Peers and Experience Drive Productivity Among Teachers

Romaine A. Campbell
Cornell University

Seth Gershenson
American University and IZA

Constance A. Lindsay
University of North Carolina at Chapel Hill

Nicholas W. Papageorge
Johns Hopkins University, IZA and NBER

Jessica H. Rendon
American University

DECEMBER 2024

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

On-the-Job Learning: How Peers and Experience Drive Productivity Among Teachers*

Workers learn on the job from both repetition and peers. Less understood is how specific types of experience and peer characteristics affect on-the-job learning. This likely differs by context (e.g., occupation, tasks, or roles). Absent such knowledge, it is unclear how to optimally assign workers to tasks and peers. We examine on-the-job learning among elementary school teachers. We focus on white teachers' productivity teaching Black students. We examine specific types of experience and specific types of peers that could lead to rapid productivity gains for white teachers: experience teaching Black students and having Black colleagues. Both lead to significant productivity gains over and above those associated with total teaching experience and access to generally productive peers. This is due to learning, as peer effects are persistent and driven by more effective Black peers. These findings offer insights to improving Black students' educational outcomes when facing a disproportionately white teaching force. More generally, they underscore the importance of understanding whether and how nuanced types of experiences and peers enter the production function and drive on-the-job human capital accumulation.

JEL Classification: I2, J24, D2

Keywords: peer effects, knowledge spillovers, teacher effectiveness, teacher diversity, achievement gaps, education production function, learning-by-doing, human capital

Corresponding author:
Seth Gershenson
School of Public Affairs
American University
4400 Massachusetts Ave. NW
Washington, DC 20016
USA

E-mail: seth.gershenson@american.edu

* Campbell gratefully acknowledges financial support from the National Science Foundation under Grant No. DGE1745303. The authors thank Stefanie DeLuca and conference and seminar participants at the University of Pennsylvania, Brown University, Auburn University, Arizona State University, American Economic Association (AEA), Association for Public Policy Analysis & Management (APPAM), Eastern Economic Association, Society for Research on Educational Effectiveness (SREE), and the Association for Education Finance & Policy (AEFP) for providing helpful feedback. Kristina Curley and Sarah Sullivan provided able research assistance. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A210434 to American University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

1. Introduction

A vast literature on human capital accumulation studies on-the-job learning, i.e., how workers become more productive over the course of their careers (Becker 1962). Outside of formal training, two mechanisms are dominant sources of on-the-job productivity increases. The first is learning-by-doing (Arrow 1962; Ben-Porath 1967). The second is peer effects, which in high-skilled occupations, are driven by knowledge spillovers (Cornelissen et al. 2017). Of course, learning-by-doing and peer effects can occur in tandem and are not mutually exclusive avenues to improving over the course of one's career (Foster & Rosenzweig 1995; Young 1993).

While these two channels describe what drives productivity gains, much less research examines how *specific* types of experience and peers may have wildly different impacts on productivity or affect different dimensions of productivity, depending on the work context (e.g., occupation, role, and tasks). Absent this knowledge, it is difficult to understand how to allocate workers to tasks or peers in a way that efficiently leverages on-the-job learning opportunities.

The current study examines how specific types of experience and peers affect teacher productivity. We examine a particular context: how white teachers improve their ability to teach Black students over the course of their careers. How Black teachers influence their white peers is an important question in its own right, as Black students benefit when they have a Black teacher. However, the teaching force is disproportionately white, so many Black children will fail to experience the benefits of having a same-race teacher (Gershenson et al. 2021, 2022). Achieving a representative teaching force is a potential long-run solution, but in the short-run it is important to learn how to improve white teachers' effectiveness in diverse classrooms.

The context we examine is ideal for examining multi-dimensional on-the-job learning since it is straightforward to measure productivity, different types of experience, and different

types of peers in schools. Using quantitative administrative data from North Carolina, we measure productivity gains as changes in teachers' impacts on students' standardized test scores, attendance rates, and suspensions. To investigate the impact of context-specific experience, we consider how long white teachers have taught Black students. To study peer effects, we examine the impact of having Black colleagues, who may possess specific insights, experience or knowledge from which white teachers can learn to more effectively teach Black children, on white teachers' Black students' achievement. We also estimate whether these two modes of learning interact. Finally, we cross-validate our findings with qualitative data collected from semi-structured interviews of teachers in North Carolina.

First, we show that both general and context-specific experience matter. Importantly, our results highlight that context-specific experience extends beyond grade-specific experience (Ost 2014) to the racial composition of classrooms: Black students perform better when assigned to white teachers with more experience teaching in racially diverse classrooms. Holding total teaching experience constant, the math scores of white teachers' Black students increase by an additional 1 to 2% of a standard deviation when their teacher has at least one year of experience in a classroom at least as racially diverse as the current classroom.

Turning to peer effects, our results show that, overall, having a Black peer increases white teachers' effectiveness with Black students, increasing math and reading scores by 2% and 1% of a standard deviation (SD), respectively. However, analyses of the full sample mask important variation by teachers' experience: novice white teachers (and their Black students) benefit greatly when they have a Black same-grade peer. Specifically, the math and reading scores of Black students of novice white teachers' increase by 6% and 5% of a SD, respectively, when their teacher has a same-grade Black peer. That these peer effects are stronger among

novice teachers is consistent with prior research on teacher peer effects (Jackson & Bruegmann 2009; Maturana & Nickerson 2019) and suggests that peer learning is the primary mechanism through which these effects operate.

Further support for this interpretation comes from the finding that the effects of historical exposure to peers of color remain statistically significant and larger in magnitude than contemporaneous exposure. Falsification tests showing that there is neither an effect of having a Black peer on white teachers' white students' achievement nor an effect of future Black peers on current outcomes bolster a causal interpretation of our results. We augment our quantitative analyses with qualitative analyses of open-ended, in-depth interviews of teachers, which largely affirm the quantitative results. The qualitative data suggest that teachers learn both formally and informally from their peers as well as through formative student interactions that enable teachers to learn about race and other elements of student identity writ large.[6]

Finally, these two channels interact in the sense that multiple teachers mentioned that difficult encounters with students in one year motivated them to seek advice from a peer, which then increased their proficiency in subsequent years. The quantitative data too finds suggestive evidence of complementarities, suggesting that peers and context-specific experience are both important sources of on-the-job learning. This finding has further implications for the allocation of resources. Moreover, it raises concerns about omitted variables bias when one studies each channel separately. For example, if the focus is on specific peer types and peer type is correlated with omitted types of specific experience, peer effects will be upwardly biased.

Our study contributes to two main literatures, one on production theory and workforce teams along and another on teacher effectiveness. We also contribute to a nascent literature that

[6] See Lindsay et al. (2024) for further analysis of the qualitative data.

considers the intersection between the two, which examines teacher peer effects and *how* teachers improve over time.

Studies on workface teams typically find that diverse-in-ability teams are more productive (e.g., Hamilton et al. 2003).[7] It is theoretically ambiguous, however, whether fielding racially and ethnically diverse teams is beneficial, as higher communication costs may hinder performance. Hamilton et al. (2012) find no effect of changing the demographic composition of teams in a garment plant on their productivity or likelihood of dissolving (holding ability constant). The related literature on workplace peer effects generally finds that having more productive peers increases individuals' productivity (Cornelissen et al. 2017). More recently, Herkenhoff et al. (2024) show that as much as 67% of workers' productivity growth "on the job" results from learning from productive coworkers. Moreover, having less productive peers does not have a symmetrically negative impact, suggesting that failure to distribute productive employees prevents the optimal development of human capital.

However, while worker productivity is likely orthogonal to race and ethnicity in a garment plant, this is not necessarily so in schools, where education is co-produced by students and teachers and teachers vary in their ability to teach students of different backgrounds (e.g., Dee 2004; Delgado 2023; Gershenson et al. 2021, 2022). Indeed, we do find that peer racial diversity matters in the context of elementary school instruction, which underscores that returns to specific peer types are heterogeneous and depend on the work context. More generally, these differences in findings underscore the importance of measuring heterogeneity in human capital accumulation via on-the-job learning across work contexts, i.e., on a case-by-case basis.

[7] This echoes the teacher literature, which finds that less effective and novice teachers benefit from effective peers and mentors, as measured by value-added (Jackson & Bruegmann 2009; Goldhaber et al. 2020; Papay et al. 2020).

We also contribute to the large literature on teachers (e.g., Chetty et al. 2014 a, b; Jackson 2019). Murnane and Phillips (1981) speculated that teachers improve over time via "learning by doing" and a large literature has since documented the returns to teaching experience (Wiswall 2013; Papay & Kraft 2015). A nascent literature investigates *why* those returns exist. Ost (2014) distinguishes between general and specific human capital, in the sense of Becker (1962), by identifying returns to grade-specific experience. Similarly, Master et al. (2016) find that teachers who taught 6 or more English language learner (ELL) students in the previous year are more effective teachers of ELL students in the current year. We add to this by showing that experience is not a monolith; rather, different types of experience develop different skills.

Similarly, Bartanen et al. (2023) identify classroom management as an important general teaching skill that improves over time, which depends on building trusting and meaningful relationships with students. This is something that novice white teachers sometimes struggle to do with students of color, as they often misinterpret their behavior and needs (Kunesh & Noltemeyer 2019; Milner & Ford 2007); our results suggest this is something that can be learned with experience. Importantly, we demonstrate that the basic idea of learning-by-doing extends to task-specific experience. The returns to teaching experience are also larger in more supportive school environments (Kraft & Papay 2014), which indicate the presence of peer effects.

Jackson and Bruegmann (2009) find significant and long-lasting increases in teacher effectiveness when they are exposed to effective same-grade peers (as measured by value-added scores). These teacher peer effects are primarily due to information transmission, or knowledge spillovers, because the effects are most apparent among novice teachers and are persistent over time. Maturana and Nickerson (2019) find that teachers, especially younger ones, are more likely to refinance their mortgage when their colleagues do. Similarly, when an effective teacher

changes schools, she increases the effectiveness of her new colleagues (Sun et al. 2017). However, the authors find no analogous harm caused by moves of ineffective teachers, which, like Herkenhoff et al., suggests the potential for Pareto-improving teacher reassignments. We expand on this work by showing that there is no single measure of what makes for an effective peer, as there are many types of knowledge spillovers and different employees may need different types of knowledge. Specifically, we identify the transmission of race-specific teaching skills (e.g., Delgado 2023; Gershenson et al. 2022) from Black peers.

Our finding that white teachers benefit from having Black peers also relates to empirical tests of Allport's (1954) "contact hypothesis," which suggests that intergroup contact can reduce racial biases (e.g., Billings et al. 2021; Carrell et al. 2019; Rao et al. 2019). White teachers' biases towards Black students are well documented (Gershenson et al. 2016; Papageorge et al. 2020) and exposure to peers of color generally, and effective ones specifically, may assuage such biases. White teachers could benefit from Black peers in a more intentional fashion as well: either by passively observing how their peers of color communicate with students of color in certain situations or by actively seeking their guidance.[8]

Finally, the current study has important implications for the efficient allocation of teachers. An obvious policy response to the increasingly diverse U.S. student population would be to create a more representative teaching force, but doing so will take time (Gershenson et al. 2021). In the meantime, policymakers and school leaders must ensure that the current teaching force, which is disproportionately white, is equipped to effectively instruct and nurture the talent in our nation's increasingly diverse classrooms. Our results provide crucial insight into this

[8] E.g., culturally relevant pedagogy (CRP) refers to the teaching practices and techniques used by many teachers of color to connect with students whose backgrounds, cultures, and lived experiences are often excluded from the mainstream school curriculum (Ladson-Billings 2008, 2022). White teachers could learn CRP from their colleagues.

problem and suggest that until the teaching workforce is truly representative of the student population in their charge, a second-best alternative may involve the strategic placement of minority teachers, accounting for both the composition and prior teaching histories of their peers.

Doing so effectively leverages the ideas that teachers' effectiveness is a function of both who and how much they have previously taught *and* that Black teachers' positive impacts on Black students not only include direct effects, but also indirect "knowledge spillover" effects through their interactions with white teachers tasked with teaching Black students. Thoughtful placement of Black teachers to increase their interactions with white teachers can thus increase the total benefits that Black teachers can have on Black students and improve the overall effectiveness of the white teaching force. And these peer placements may take precedence in cases where white teachers have had little prior experience teaching in diverse classrooms. Outside of our specific context of teachers, in any occupation or workplace, there likely are unique types of experience and peers that increase productivity. This illustrates the need to better understand the nature of industry-specific production functions and the way that specific types of experience or peers enter those production functions when considering how to allocate workers.

The paper proceeds as follows: Section 2 describes the quantitative administrative data. Section 3 describes the identification strategy. Sections 4 and 5 present the quantitative results, for the returns to experience and peer effects, respectively. Section 6 describes the qualitative analysis. Section 7 concludes.

2. Administrative Data from North Carolina

We analyze administrative data from the North Carolina Education Research Data Center (NCERDC). In partnership with the North Carolina Department of Public Instruction, the

NCERDC collects data on all public-school students in the state, including district-, school-, and teacher-level data. These data are publicly available to researchers who pay a usage fee and satisfy certain data security requirements (Gershenson & Langbein 2015; Muschkin, Bonneau, & Dodge, 2011). These are the same data analyzed by Jackson and Bruegmann (2009).

Summary statistics for our analytic sample are presented in Tables 1 and 2. Table 1 summarizes the data at the student-year level, which is the unit of analysis in our regression models, and Table 2 does so at the teacher-year level, which is the level at which treatment varies (i.e., teaching experience and peer quality). In Table 1, column 1 summarizes all students while column 4 does so for our primary analytic sample of 307,986 student-year observations for Black students matched to white self-contained classroom teachers in grades 3 through 5 between 1995 and 2018 in North Carolina public schools. The main outcomes are end-of-grade state test scores for math and English Language Arts (ELA). Test scores are standardized across all students in the NCERDC data to have mean zero and unit variance within grade-year. The students in our main analytical sample have an average standardized math score of -0.41 with a standard deviation of 0.88. They have an average standardized ELA score of -0.37 with a standard deviation of 0.91. The negative averages are consistent with previous findings in the literature documenting Black students' underperformance relative to their white peers.

We also consider non-academic outcomes that teachers are known to influence, namely attendance and suspensions (Gershenson 2016; Jackson 2019; Lindsay & Hart 2017; Liu & Loeb 2021). Students in our main analytic sample are absent on average 4.7 days per academic year with a standard deviation of 6.55. Approximately 5 percent of students are considered chronically absent (i.e., absent for at least 18 days in an academic year). Students have on

average 0.25 out-of-school suspension days with a standard deviation of 1.44. Each year, about 7 percent of students have ever received out-of-school suspension.

These student-year observations map to 23,908 teacher-year observations, summarized in column 4 of Table 2. By design, the main analytic sample contains white teachers who have at least one Black student. Here, the average class size is around 18 students with a standard deviation of 6.5 and classrooms, on average, contain 44 percent Black students. About 29 percent of teachers hold an advanced degree. Most teachers in our sample (81 percent) have received a regular state license instead of, for example, working under a provisional or temporary license and about 10 percent of teachers have national board certification.

We standardize teachers' license exam scores on the elementary and early childhood education tests required for all North Carolina teachers to have mean zero and unit variance in each year. Teachers in our main analytical sample have an average score of 0.15 with a standard deviation of 0.67. Approximately 89 percent of teachers are female. Teachers have average value-added of 0.03 and 0.02 with standard deviations of 0.48 and 0.43 for math and ELA, respectively.

2.1 Peer Characteristics

Approximately 40 percent of white teachers have a Black peer in their same grade in a given academic year. This extensive margin indicator will be our primary treatment, though we will also consider an intensive margin measure of the share of peers who are Black. The 40% figure is relatively unchanged when we consider both Black and Hispanic peers, because there are relatively few Hispanic teachers in our data. Following Jackson and Bruegmann, we also report the average value-added measures in math and ELA of teacher's same-grade peers, along

with other observed qualifications, in Table 2. Teacher value-added is estimated using an adjusted test score growth model using data from 1995-2000, i.e., from out of sample.

2.2 Experience

Approximately 9 percent of teachers in our sample are new to teaching. Conditional on not being new, the average teacher in our analytic sample has 10.6 years of teaching experience in North Carolina. However, as documented in Ost (2014), this aggregate measure overlooks important nuance in different types of experience. In terms of same-grade experience, we see that, each year, about 29 percent of teachers in our sample are new to teaching a specific grade. Conditional on not being new, the average teacher has about 4 years of teaching experience in the grade. A key innovation in the current study is to consider experience teaching in racially diverse classrooms—a variable which we name similar. Motivated by Master et al. 2016., for each class, similar is defined as the number of prior years (excluding the current one) in which the teacher has had a class with *at least* the same share of Black students (rounded to the nearest decile) as the current class. In terms of similar experience, each year, about 37 percent of teachers are new to teaching a diverse class. Conditional on not being new, the average teacher has about 3.4 years of experience teaching similarly diverse classrooms.

3. Identification Strategy

Generally, we augment commonly used lag-score value-added models of the education production function to include more nuanced measures of teachers' teaching experience and peer characteristics. Specifically, we build on the models and identification strategies presented in Ost

(2014) and Jackson and Bruegmann (2009) which we outline in sections 3.1 and 3.2, respectively. Sensitivity analyses and robustness checks are described in section 3.3.

3.1 Identifying Returns to Experience

Ost (2014) made an important contribution to our understanding of how and why teachers improve over time by differentiating between a teacher's *total* years of teaching experience (a measure of general human capital) and their experience teaching in the current grade (a measure of specific human capital). Ost identified the additional benefit of same-grade experience by distinguishing between the two types of experience in models that condition on teacher fixed effects (FE) and year-by-grade FE. We add a third type of experience to similar models, which we term experience in racially similar classrooms and focus on the outcomes of Black students of white teachers. This is another form of specific human capital distinct from grade-specific teaching experience.

Following Ost, we specify nonparametric specifications of all three experience types and estimate linear models of the form:

$$y_{ijgst} = \beta X_{it} + \gamma W_{jt} + \theta_{gt} + \omega_{st} + \varphi_{js} + f(yrs) + g(yrs_same) + h(yrs_similar) + u_{ijgst}, \quad (1)$$

where y is a year-specific outcome such as standardized EOG test scores, suspensions, or absences; X is a vector of observed student characteristics that includes socio-demographic controls and lagged math and reading scores; and W is a vector of observed classroom characteristics that includes class size and observed teacher characteristics such as certification status, educational attainment, and demographic controls. The θ, ω, and φ represent grade-by-year, school-by-year, and teacher-by-school FE, respectively. Finally, f, g, and h represent general functions of each type of experience. Ultimately, equation (1) depicts a standard value-

added specification that controls for lagged achievement, which the literature agrees sufficiently adjusts for nonrandom sorting of students to classrooms (Chetty et al. 2014 a). The various fixed effects further control for sorting into schools and differences across years in the EOG tests.

We estimate the multi-dimensional FE model specified in equation (1) using the linear `reghdfe` Stata package and estimator (Correia 2016). We also estimate Poisson regression analogs to equation (1) for count outcomes such as absences and suspensions (Correia et al. 2020). We cluster standard errors by teacher-year, as this is the level at which the treatment of interest varies and all students of a given teacher, in a given year, receive the same treatment (Abadie et al. 2023); that said, in the appendix we show that the main results are robust to clustering at higher levels such as the teacher, school-year, and school, as well as to two-way clustering by student and teacher-year, teacher, school-year, or school (Cameron et al. 2011).

3.2 Identifying Peer Effects

To understand how Black peers affect white teachers' effectiveness educating Black students, we augment the regression models estimated by Jackson and Bruegmann (2009) to include a measure of peer race in the vector of peer characteristics as an input in the education production function. Intuitively, the identification strategy pioneered by Jackson and Bruegmann includes the means of observable characteristics of teachers' current, same-grade peers (e.g., certification status, experience, and value-added measures of effectiveness) as additional measures of teacher quality in value-added models of the education production function.

The preferred specification introduces peer race via a binary indicator for having at least one (current) Black colleague. However, we consider two alternative specifications. First, in the spirit of taking the mean characteristics of one's peers, we include the share of current peers who

are Black. Second, given that Jackson and Bruegmann (2009) find that peer quality, as measured by experience and test score value-added, we similarly distinguish between having a more or less effective Black peer using a set of mutually exclusive indicators. Importantly, we continue to control for all three types of teaching experience in these models, though for parsimony we replace the nonparametric specification with a continuous measure of experience and an indicator for being "new."

Specifically, we estimate models of the form:

$$y_{ijgst} = \beta X_{it} + \gamma W_{jt} + \theta_{gt} + \omega_{st} + \varphi_{js} + f(Peer) + u_{ijgst}, \quad (2)$$

where the notation is the same as in equation (1), with two exceptions: the vector W now includes all three types of teacher j's year-t experience and we have added the aforementioned vector of the teacher's school-grade-year peers' average characteristics ($Peer$). As with equation (1), we will estimate equation (2) using the full sample as well as the sample of white teachers' Black students. We will further stratify the data to estimate equation (2) separately by teachers' experience level, as relatively novice teachers may benefit more from the presence of a high-quality peer. The baseline estimates will define novice as two or fewer years of experience, though we will verify that the results are robust to this arbitrary decision.

The main innovation in Equation (2) is the inclusion and specification of the *Peer* vector, which contains objective measures of teacher j's same-grade peers in year t. This vector includes the aforementioned peer-race variable in addition to the measures of peer quality used in Jackson and Bruegmann (2009): the average of observable characteristics like experience, certification status, and education, and the average of the peers' estimated value-added. Teacher value-added is estimated using an adjusted test score growth model using data from 1995-2000. We use these pre-sample estimates of value-added when investigating the effect of peer quality on student

achievement using data from 2001-2018. Because these value-added estimates are time-invariant, any variation in mean peer value-added is due to changes in the composition of a teacher's peers. The disadvantage to this approach is that teachers who are not in the pre-sample data (1995-2000) will not have a value-added estimate. Following Jackson and Bruegmann (2009), we still use the full sample of teachers by using mean imputation for teachers with missing value-added estimates and including an indicator for missing value-added.

3.3 Sensitivity Analyses

Adjusting for fixed effects at the school-year level will ensure that we compare student outcomes of different teachers in the same school, in the same year. We control for Black colleagues directly affecting students in other classrooms by exploiting within-teacher variation over time in exposure to Black colleagues via the teacher-by-school FE. We also control for grade (or grade-by-year) indicators to flexibly account for differences across grades. Hence the key identifying assumption when estimating equations (1) and (2) is that in a given year, Black teachers in the school are (conditional on some basic teacher and student controls) randomly distributed across grade levels.

We probe the plausibility of the identifying assumption using two falsification tests. First, unique to our context, we show that having a Black peer does not affect the performance of white teachers' white students or Black teachers' Black students. This suggests that the documented effect is the transmission of some form of racial competency and not a more general teaching skill uniquely possessed by Black teachers. Second, as in Jackson and Bruegmann, we show that in equation (2) leads of the *Peer* variables (the characteristics of future peers) do not affect current performance.

4. Learning by Doing

In this section, we follow Ost (2014) in two notable respects. First, we restrict the analytic sample to new teacher cohorts, in an effort to observe their entire teaching history in North Carolina's public schools, and therefore have accurate records of their specific teaching experiences. Second, we employ nonparametric specifications of both total and grade-specific teaching experience, where no prior experience is the omitted reference category. Table 3 reports our baseline estimates of equation (1) that document the impact of teaching experience on standardized end-of-grade (EOG) math scores. Appendix Table A1 reports analogous results for EOG reading scores, where effects follow similar patterns but are smaller in magnitude, consistent with the general finding in the economics of education literature that school-based inputs have larger impacts on math scores, perhaps because ELA skills are more often practiced at home (Currie & Thomas 2001).

Columns 1 and 2 of Table 3 replicate Ost's main results in the full sample of all students in all self-contained classrooms. Column 1 shows a large, statistically significant increase in effectiveness of about 4% of a test-score SD in the first year of teaching and improvements of about 1 to 2% of a test-score SD thereafter. This is consistent with the general finding of returns to teaching experience that are particularly large in the first year (Ost 2014; Wiswall 2013). In column 2, we add indicators for years of grade-specific teaching experience. These indicators are individually and jointly statistically significant. Consistent with Ost, these estimates imply that (i) failing to account for whether or not teaching experience occurred in the grade currently being taught understates the returns to experience and (ii) experience in the same grade increases the return to general experience by about 50%. That we replicate the main results of Ost (2014)

using a larger and more recent sample (2001-2018 rather than 1997-2012 cohorts) and a slightly different model specification suggests that this is a robust finding.

Columns 3 and 4 of Table 3 replicate the grade-specific results from columns 1 and 2 in our primary sample of interest: Black students of white teachers. The returns to general and grade-specific experience for this subset of teachers are again strongly statistically significant, and similar in magnitude, to those for the full sample. Finally, columns 5 and 6 augment the model to include our new variable of interest, which we call "years similar" in the table. This is a measure of teachers' experience in classrooms with at least as great a share of Black students. Column 5 adds this nuanced experience measure alongside total years of experience while column 6 includes it alongside both total and grade-specific experience. In both cases, the race-specific experience indicators are jointly and individually statistically significant.[9] This suggests that there is another dimension to which the specificity of teaching experience matters: who is in the classroom. For white teachers, holding total experience constant, their Black students' EOG math scores increase by about 1 to 2% of a test-score SD when their teacher has had at least one year of experience in a classroom that was at least as racially diverse as the current classroom.

Table 4 re-estimates the full model with all three types of experience shown in column 6 of Table 3 for five additional educational outcomes. Column 1 repeats the math EOG results as a reference point. Column 2 reports EOG reading results. The effects of total experience on reading scores are about half as large as those on math and less precisely estimated; this is consistent with other studies on the returns to experience, the returns teacher quality, and the impacts of other school-provided inputs more generally (e.g., Gershenson 2016). The returns to

[9] Statistical inference in this and all tables presented in the main text is based on standard errors clustered at the teacher-year level. Appendix Table A2 shows that the estimates in Table 3 are (i) robust to using a continuous measure of experience and (ii) remain statistically significant when clustering standard errors at different levels.

grade-specific experience on reading EOG scores are indistinguishable from zero in the sample of white teachers' Black students, but they are significant and in line with Ost's (2014) results in the full sample (See Appendix Table A1). However, like in the case of EOG math scores, we do observe significant effects of experience in racially diverse classrooms on Black students' EOG reading scores. These effects on reading EOG scores are slightly smaller, but similar in magnitude and jointly significant.

Columns 3 and 4 of Table 4 examine the effect of teacher experience on absences and chronic absences rates, respectively, where chronic absence is defined as being absent for at least 18 days (10% of school days). Consistent with existing evidence from North Carolina (Gershenson 2016; Ladd & Sorensen 2017), total experience significantly reduces student absenteeism and rates of chronic absenteeism, though grade-specific experience attenuates this effect. This could be the result of teachers with less grade-specific experience focusing more on socioemotional skills; however, this finding falls outside the scope of the current study and so we leave it to future work. Finally, while the coefficients on the experience with similarly racially diverse classroom indicators tend to be negative, which suggests that white teachers with such experience reduce their Black students' absences, they are imprecisely estimated.

Columns 5 and 6 of Table 4 examine the effect of teacher experience on the count of out-of-school suspension (OSS) days and whether the student was ever suspended during the school year, respectively. This analysis is motivated by the fact that teacher race, and specifically student-teacher race match, is a known predictor of exclusionary discipline (Hayes et al. 2023; Holt & Gershenson 2019; Lindsay & Hart 2017). Here, we see that prior experience in racially diverse classrooms significantly reduces the number of OSS days and marginally reduces the

likelihood of ever receiving an OSS for Black students of white teachers.[10] In sum, the results presented in Tables 3 and 4 reaffirm the ideas that (i) effective teaching is multidimensional and teachers affect both cognitive and non-cognitive outcomes (e.g., Jackson 2019), (ii) teachers improve over the course of their careers and that the context in which experience is accrued matters, and (iii) white teachers' capacity to effectively educate students from other backgrounds improves over time, and even more so with repeated exposure to diverse classrooms.

Finally, Table 5 estimates the baseline experience regression from column 6 of Table 3 for a few different samples. Recall the main analytic sample contains only the Black students of white teachers, as the hypothesis is that white teachers "learn by doing" such that their effectiveness educating Black students increases with repeated exposure to Black students. In Table 5, we estimate the model for all teachers of all students in column 1, for white students of white teachers in column 2, and for Black students of Black teachers in column 3; column 4 reproduces the baseline results for comparison. This exercise is motivated by two related, but distinct, ideas. First, experience teaching in classrooms with different demographic compositions might contribute to the development of teaching, communication, and classroom management skills that are orthogonal to racial competency.[11] Thus, if we observe effects of prior experience teaching Black students on the achievement of white teachers' white *students*, it suggests another mechanism through which this type of experience benefits teachers. Second, viewed as a falsification test, if we observe white students benefitting *more* than Black students from their white teachers' experience in diverse classrooms, this will cast doubt on our identification strategy since it is hard to come up with a story why this would be so.

[10] Appendix Table A3, Panel A, reports Poisson regression analogs to columns 3 and 5 of Table 4. The Poisson estimates are qualitatively similar, suggesting that the results are robust to the use of a linear model.
[11] Alternatively, it could represent a spillover effect: becoming more effective with Black students could free up bandwidth that enables them to be more effective with white students too.

The results in Table 5 emphatically support the idea that white teachers' exposure to diverse classrooms benefits *all* their subsequent students, though the benefits to their Black students are about twice as large as those accruing to their white students, as seen by comparing the estimates in column 2 to column 4. This suggests that white teachers do in fact learn a variety of broadly applicable skills from their experiences teaching in diverse classrooms, but also that at least some of those skills are unique to (or more impactful for) nonwhite students. Equally interesting are the results in column 3, which show that experience in classrooms with Black students is the most important type of experience that Black teachers can accrue.[12] In fact, these effects are similar in size to those in the baseline sample of white teachers' Black students, suggesting that teachers of all backgrounds benefit from experience in diverse classrooms, particularly when it comes to effectively educating Black students.

5. Learning from Peers

5.1 Main Results

Having seen that teachers improve along multiple dimensions over the course of their careers and that said improvements are moderated by the grade levels and classroom composition in which that experience was accrued, we now turn our attention to the role that peers may play in helping teachers improve upon multiple dimensions over the course of their careers. As described in section 3, we build on the analysis of Jackson and Bruegmann (2009), which identifies teacher peers at the school-grade-year level and measures peer quality as the

[12] The insignificant returns to Black teachers' general teaching experience is interesting in its own right and suggests that Black teachers accumulate human capital differently from white teachers, though we leave further investigation of this to subsequent research.

average of observed peer characteristics as well as the average math or reading value-added measures (VAMs) of the peers. We add to this an indicator of whether teachers had a Black peer.

Table 6 reports our baseline estimates of equation (2) using math EOG scores as the outcome.[13] Panel A uses samples of all students and teachers of all racial backgrounds. Column 1 replicates the basic result of Jackson and Bruegmann (2009) in our full dataset: having peers whose average math VAM is 1 SD higher increases students' current math EOG scores by about 4% of a test-score SD. This is about twice as large as the effect of peers' reading VAM scores on reading EOG scores, as shown in Appendix Table A4. In column 2 we augment this model to include an indicator for having at least one Black peer. The coefficient on this indicator is positive but small and imprecisely estimated.

Following Jackson and Bruegmann (2009), in columns 3 and 4 of Table 6 we split the sample by novice and veteran status, respectively, as more novice teachers are likely more receptive to, and in need of, peer feedback. As expected, and consistent with Jackson and Bruegmann (2009), the effect of peers' math VAM is 50% larger for novices than for veterans, though it is significant for both groups. Similarly, the effect of having at least one Black peer is an order of magnitude larger, and becomes statistically significant, for novice teachers. Novice is defined as having 2 or fewer years of experience, as our earlier experience results indicate that teachers improve markedly in their first couple of years; that said, Appendix Table A5 replicates column 3 of Table 6 using various definitions of novice and shows this result is robust to the choice of cutoff. Appendix Table A6 reproduces Table 6 clustering the standard errors at different levels and verifies that the results are robust to this decision.

[13] Appendix Table A4 presents an analogous version for reading EOG scores.

Panel B of Table 6 estimates the same four regressions using the sample of Black students of white teachers. Column 1 shows that peers' average math VAM is slightly more important in this sample. Column 2 shows the black-peer indicator's effect is larger and more precisely estimated than in the full sample: when a white teacher has a Black peer, their Black students' math EOG scores increase by 1.6% of a test-score SD, which amounts to a 4% increase. For novice white teachers of Black students, column 3 of Panel B shows that having a Black peer boosts Black students' math performance by 5.7% of a SD, which amounts to an 11.6% increase. This is compared to markedly smaller, and imprecisely estimated, effects amongst veteran teachers.[14]

Finally, Panels C and D of Table 6 differentiate the quality of Black peer that white teachers encountered. If novice teachers are more affected by peers because they are less experienced and more open to learning, a similar logic suggests that said learning is greater when the Black peer is themselves more experienced or more effective. Accordingly, in panel C we dichotomize the any Black peer indicator into two mutually exclusive indicators for having a veteran (as opposed to novice) Black Peer while novice is again defined as having two or fewer years of experience.[15]

The results in panel C clearly show that who the Black peer is matters: the effect on novice white teachers is entirely driven by veteran Black peers with 3 or more years of experience. This is likely due to the combination of more experienced Black peers being more

[14] Importantly, these results are also robust to how the Black peer treatment is defined. This is shown in Appendix Table A7, which replaces the binary indicator for having at least one Black peer with the percent of peers who are Black, and finds qualitatively similar results that are consistent with no dosage effects of Black peers: shifting from 0 to 100% increases white teachers' Black students' math EOG scores by 3% of a SD and by 7% of a SD for Black students of white novice teachers, though these effects are imprecisely estimated. We prefer the binary (extensive margin) definition of treatment used in the main text because peer groups vary in size and there are likely diminishing returns to having multiple good peers.

[15] We cannot include all four indicators in the same model due to overlap in the veteran and effective Black peer indicators.

comfortable or confident in their advice, and the white teachers being more receptive to advice from Black peers who are more experienced.

Panel D similarly distinguishes between more and less effective Black peers, where no Black peer is again the omitted reference group. "More effective" is defined as having a math VAM > 1, or one SD above the average. The effect of having a Black peer on novice white teachers is more than twice as large when the peer highly effective, as measured by their math VAM. Moreover, highly effective Black peers improve the performance of veteran white teachers by a significant amount as well. Interestingly, the effect of average math VAM of all peers remains significant and of a similar magnitude in these regressions, again suggesting that different types of peer learning are occurring from different peers.

Table 7 re-estimates the same peer-effect models for the Black students of white teachers for a variety of other educational outcomes. The first 3 columns of Panel A replicate the math EOG results and columns 4 through 6 do so for reading EOG. The effect of having at least one Black peer on novice white teachers' Black students' reading EOG scores is 4.5% of a SD, which resembles that on math EOG scores and represents a 10% increase.

Panels B and C of Table 7 investigate peer effects on student absences and suspensions, respectively.[16] Having a Black peer significantly reduces novice white teachers' Black students' absences. Column 2 of Panel B shows a reduction of about two thirds of an absence, or a 13% reduction, while column 5 shows a 2.3 percentage point reduction in the chronic absence rate, which is nearly a 50% reduction. Similarly, columns 2 and 5 of Panel C show similar effects of having a Black peer on novice white teachers' Black students' exposure to exclusionary discipline. For example, on the extensive margin, column 5 shows a 2 percentage point (25%)

[16] Panel B of Appendix Table A3 replicates the absence and OSS-days results using a Poisson regression model.

reduction in the likelihood of ever being suspended from school. Together, Tables 6 and 7 reaffirm the idea that teachers learn from their peers along multiple dimensions in ways that enable them to improve students' academic and socioemotional outcomes and – specifically – that they learn different skills from different types of peers.

5.2 Falsification Exercises

Table 8 estimates the baseline peer-effects regressions from columns 2-4 of Table 6 for two different samples: Black students of Black teachers and white students of white teachers. If the results observed in Tables 6 and 7 are driven by Black peers sharing general insights about racial competency or providing advice regarding specific students or incidents involving race with their white colleagues, then the Black-peer effect should not appear in either of these samples. It does not: the Black-peer indicator is indistinguishable from zero in each column of Table 8. In the white-student white-teacher sample (columns 4-6), the estimates are fairly precisely estimated zeroes. This reinforces our interpretation of the peer effect finding and its validity: Black peers increase novice white teachers' capacity to effectively teach Black students.

Finally, recall that we hypothesized two key mechanisms that help to explain the effect of white teachers' exposure to Black same-grade peers on the outcomes of their Black students. The first mechanism involves intergroup contact, which has the potential to reduce racial biases. The second mechanism pertains to peer learning, which can enhance white teachers' effectiveness in educating students of color. To validate these mechanisms and our research design, we now follow Jackson and Bruegmann (2009) in estimating a series of models that include lags and/or leads of the Black-peer indicator. If lagged exposures are significant, this indicates learning that

sticks with teachers. If leads (future exposures) are significant, this would indicate nonrandom sorting that our fixed effects fail to control for.

The leads-and-lags model estimates are reported in Table 9, again restricting the sample to Black students of white teachers. Because adding leads and/or lags requires multiple years of data we cannot include rookie teachers in the sample, nor can we fruitfully use the novice distinction made in previous tables. Accordingly, Panel A reports the baseline model (with no leads or lags) estimated using the analogous leads / lags samples. Panel B reports the actual leads / lags model estimates. Columns 1 and 2 add one and two lags to the baseline model, respectively. In neither case does adding the lag change the estimated coefficient on the current Black-peer or peer math VAM inputs, though the lagged peer math VAM measures are significant, suggesting a lasting impact of having an effective peer. The first lag of having a Black peer is statistically insignificant, but the second lag increases Black students' math EOG scores by about 2.4% of a SD, which is nearly identical to the contemporaneous effect of a Black peer for novice teachers observed in Table 6. This stickiness suggests that at least part of the effect of having a Black peer is due to learning, as opposed to the Black peer actively intervening or causing the white teacher to alter their effort. That said, there is a mechanical result here too since teachers are necessarily younger when the twice-lagged peer was encountered, and we have already seen that peer exposures are more impactful earlier in one's career.

Columns 3 and 4 of Table 9 add one and two leads to the baseline model, respectively. Doing so does not appreciably change the estimated contemporaneous effect of having a Black peer nor are the leads ever individually or jointly statistically significant. Viewed as a falsification exercise, this lends credibility to the baseline estimates by suggesting there is no endogenous sorting into the treatment condition of having a Black peer. The same is true in

column 5, which adds one lead and one lag to the baseline model. Together with the results presented in Table 8, which showed no effect of having a Black peer for the white teachers of white students or the Black teachers of Black students, the insignificance of the leads in Table 9 suggests that the general finding that the Black students of white teachers benefit when their teacher has a Black peer can be given a causal interpretation.

6. Qualitative Analysis

6.1 Data and Method

The quantitative results presented in sections 4 and 5 provide compelling evidence that teachers learn from past interactions with students and their peers in a way that relates to racial competency, as measured by white teachers' effectiveness in teaching Black students. Here we describe our qualitative data collection and analysis, which corroborates the quantitative findings and provides some more nuanced information about how white teachers learn racial competency from their peers. A richer qualitative analysis is provided in Lindsay et al. (2024).

Specifically, we solicited semi-structured interviews from a purposive sample of public-school teachers in three racially diverse counties in North Carolina between Spring 2022 and Spring 2023. We successfully interviewed 39 teachers, of whom 27 were white, 7 were Black, and 30 were female. Their tenure in NC public schools ranged from 1 to 34 years. We first sought to determine whether white teachers report having learned from their Black peers, particularly regarding pedagogy and interaction with Black students. After observing that this did occur, we then sought to determine the common means through which white teachers report having learned from their Black peers, be it through observation, direct instruction, or some other mechanism.

We modeled our interview design after the narrative interviewing approach, a well-documented method in urban sociology that employs open-ended questions to elicit natural, long-form responses (e.g., Boyd & DeLuca 2017; DeLuca et al. 2016). By avoiding overly specific probes at the onset, narrative interviewing promotes greater coverage of topics and detail across interviews. When used in combination with encouraging verbal cues and body language, an open-ended question structure encourages the respondent to tell complete stories, without fear of judgment or failure to properly abide by a perceived interviewing protocol. As a result, narrative interviewing yields rich data sets optimal for grounded approach coding.

Coding for the purpose of this analysis designated segments related to the following: teaching approaches, struggles, goals, beliefs, and discipline styles in general; teaching approaches, struggles, goals, beliefs, and discipline styles specifically in regard to teaching students of different races; teachers having learned or developed their teaching style due to influence from students, mentors, peer teachers, and principals; opportunities for learning from teachers of a different race or lack thereof; personal definitions of equity and any influences on understanding of equity; school climate, staff relationships, school practices, and school culture; and formal versus informal learning experiences. Coding also included indicators for any discussions of teaching experiences as novices, their school's racial and socioeconomic composition, experiences during the pandemic, and experiences with parents. A "reflexivity" code encompasses segments related to a teacher demonstrating or failing to demonstrate some sort of reflection on their growth mindset and self-awareness regarding racism or racial identity.

Of the 39 respondents for whom we have complete, coded data, nearly all (36) discussed experiences teaching students of different races and almost 85% (33) reported challenges in this regard. About 85% (33) teachers discussed the development of their teaching styles and 90%

(35) attributed a nontrivial share of their pedagogical development to learning from peers. About 72% of respondents (28), including all but one of the white respondents, specifically discussed the impact of Black peer teachers on their professional development.

6.2 Experience Results

All teachers acknowledged improving over the course of their careers. About 85% (33) teachers discussed specific aspects of their teaching strategies that improved over time. About two-thirds (68%) identified specific instances that generated improvement or effort on their part to improve. For example, one teacher said:

> *"I can go back through my teaching years, and I can say: this is when this child and my experience with this child helped me learn how to do this. Or, my experience with this child helped me understand that I didn't know enough and I needed to go and find out more information about this. I would say they [prior classes / students] definitely have had a big influence [on my professional development]."*

The majority (about 65%) acknowledged that the demographic composition of a particular classroom either challenged them or caused them to seek specific assistance. For example, one teacher noted:

> *"I think my diverse learners have really impacted how I teach, knowing that they need something different… Their learning differences or their behavioral differences, they just want to be part of a class. They want to be included as much as they can, so I go out of my way to make sure that they're included and feel that they belong… I think that [that realization has] changed my thinking a lot."*

Similarly, another white female teacher explicitly acknowledged the cultural differences and lack of racial congruence between herself and many of her students:

> *"Because my early career, I started out teaching children who didn't look like me and most of those children didn't speak English either. So, there was the language barrier, there was a huge cultural barrier because schools in the countries they came from were very different than school in America. I feel like that has always been my teaching. I very rarely had classes that looked 100% like me. But I would say that it's important to incorporate that culture into your learning, whether you're inviting the families in to*

share, whether you're including in the readings and the literature that we're doing. I think that that is key to making children feel welcome and belonging in your classroom. I think that that as important, especially when your teacher doesn't look like you."

Together, the qualitative data collected and coding regarding teachers' improvement over time and specifically how prior classroom interactions influenced that improvement (and efforts to improve and seek help) reaffirm the quantitative evidence presented in section 4. It also provides a natural segue to section 6.3, as peers were a common source of help that teachers, and specifically novice teachers, sought when facing a new or challenging situation.

6.3 Peer Results

Of the 28 respondents who specifically mentioned the value of discussing teaching Black students with their Black peers, half (14) indicated that these interactions occurred during their time as a novice teacher (i.e., three of fewer years of experience). Such discussions provide evidence of "spillover" effects and confirm our quantitative findings that Black peer teachers positively influence the pedagogy and racial competence of novice white teachers, which ultimately results in improved outcomes for their Black students. For example, a white female teacher explained:

I have been really fortunate [that] out of our four 5th grade teachers, two of them are Black women and then me and the other one are White women and I have really appreciated and grown with the ability and openness within our [group], like, hey, this happened today, I just want to talk through it and make sure that I'm in my appropriate space or how can I handle this situation or here is something that a student said and I'm not sure if that needs to be addressed and putting our heads together with our different perspectives and getting to a place that's okay to [inaudible], but also to call someone out lovingly, like, you said this and that didn't make me feel okay…

Having confirmed that teachers report experiencing and benefiting from conversations about teaching Black students with Black peers, we turn to an inquiry into the mechanisms through which this spillover occurs. As mentioned previously, teachers describe how both formal

and informal interactions with their peer teachers influence their teaching practices. In our interviews, we see evidence of spillover operating through both of these channels: nearly all teachers (37 and 38, respectively) reported learning from peers through formal means, like assigned mentorship programs, and through informal means, like casual conversation and friendship. For example, one novice teacher described the following experience with her formal mentor as follows:

> *In this first year, I would say, I've had a really good mentor at my school. He's an English teacher. He teaches right next door to me. I've gone to him with probably 500 questions this past school year, so he's been influential in having that on-site support of someone that I know that I can go and ask a question to and I can feel comfortable talking about places where I think that I failed has been an excellent influence.*

She continues, discussing her positive opinion of a specific, organized peer training program:

> *Also, I would say there were a few things that we learned in MAT (Master of Arts in Teaching program) that I've thought about in my first year that have influenced my practice. We talked a lot about funds of knowledge and about understanding what students already know, trying to incorporate that and recognizing that they have strengths that they can use in the classroom. That's been a huge thing. Making sure that the talks that we've had about incorporating, especially in an English class, diverse texts because it's really easy, especially this first year of my career, it's been really easy to just stick with what's always been done because there is so much material for it and all the other teachers are teaching those texts and it's like, okay, it's just easiest to stick with that and so, I haven't done it to the extent that I want to, but trying to integrate some diverse texts.*

In terms of informal learning from peers, when the interviewer asked a novice white teacher about who she goes to for help, she described her interactions with a peer teacher of a different race:

> *My first year, it was [name – the teacher she spoke about], just going to him and being like, I need help. And just telling him, I don't know, and I don't know who to ask, and I feel stupid for asking, but I need help because I don't know X, Y, or Z. And he would help me out with whatever I needed to be. And he didn't judge me either. I want to be better. I want to know better and do better moving forward.*

This respondent finds value in interracial peer interactions, with her peer teacher's advice contributing to her goal of becoming a better teacher for her students of other races.

The quantitative analysis of administrative data from North Carolina described in section 5 provides compelling evidence of spillover effects, as the presence of at least one Black same-grade peer produces a significant increase in the test scores of Black students of white novice teachers. The qualitative data discussed in the current section allows us to confirm and to better understand the mechanisms through which these spillover effects arise and, more generally, how teachers improve their performance in diverse classrooms over the course of their careers. Evidence from the qualitative data shows how both formal and informal interactions among peer teachers of different races produce positive effects on teacher attitude and performance. These characteristics improve teacher quality, which is understood to be associated with improved student outcomes (Chetty et al. 2014 a, b; Jackson 2019). These narratives allow us to conclude that spillover effects of Black teachers on the Black students in other classrooms, specifically those of novice white teachers, are the causal result of peer learning through both formal and informal interactions with peer teachers.

7. Conclusion

This study provides evidence from North Carolina that white teachers learn how to teach Black and Hispanic students while on the job. They learn through experience and from their peers. Crucially, they learn more through specific kinds of experience (teaching in diverse classrooms) and specific types of peers (colleagues who are Black teachers). Other findings suggest that peer effects are concentrated among newer teachers and that there is a potential complementarity between relevant experience and high-quality peers.

Specific to the context we study, our findings suggest an opportunity for closing racial and ethnic achievement gaps right now by more thoughtfully assigning teachers to peers and tasks that improve their ability to teach diverse classrooms. While the benefits of having a diverse and representative teaching force are well documented, achieving such a teaching force will take time (e.g., Gershenson et al. 2021). Creating opportunities for learning to occur in the disproportionately white teaching force is therefore a potentially useful and cost-effective strategy to pursue in the meantime. Moreover, our findings suggest that the positive impacts Black teachers have on Black students is understated if we focus solely on their direct interactions with students in their classrooms and fail to appreciate indirect impacts via knowledge spillovers to white teachers. Organic and informal interactions that facilitate peer learning could also be formalized and incorporated into teacher training programs and in-service mentoring programs. In general, our findings provide practical guidance on how to use the teaching workforce we have to help close frustratingly persistent achievement gaps and, importantly, suggest that lower productivity among white teachers who teach Black students is not fixed, but could be improved given the right experiences, training, and peers.

More broadly, our findings suggest that our understanding of on-the-job learning through peers and experience is incomplete. The type of experience and the type of peers matter greatly. Earlier research has made the point that more productive peers tend to be more valuable than less productive peers and that task-specific experience drives learning and productivity gains. Our contribution is to consider these factors in a setting where both can be measured simultaneously. This way, we can show that both type of peers and type of experience matter and, moreover, the returns likely differ across different work contexts. Future research could use these findings as a basis to examine different kinds of peers and experience in a variety of settings, which would

provide more precise lessons on how to evaluate on-the-job learning and allocate human resources in a way that more efficiently allocates workers to peers and experiences.

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2023). When should you adjust standard errors for clustering?. *The Quarterly Journal of Economics*, *138*(1), 1-35.

Allport, Gordon W. (1954). *The Nature of Prejudice*. Oxford: Perseus Books Group.

Arrow, K. J. (1962). The economic implications of learning by doing. *The Review of Economic Studies*, *29*(3), 155-173.

Bartanen, B., Bell, C., James, J., Taylor, E. S., & Wyckoff, J. H. (2023). "Refining" Our Understanding of Early Career Teacher Skill Development: Evidence From Classroom Observations. Annenberg EdWorkingPaper No. 23-845.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, *70*(5, Part 2), 9-49.

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy*, *75*(4, Part 1), 352-365.

Billings, S. B., Chyn, E., & Haggag, K. (2021). The long-run effects of school racial diversity on political identity. *American Economic Review: Insights*, *3*(3), 267-284.

Boyd, M. L., & DeLuca, S. (2017). Fieldwork with In-Depth Interviews: How to Get Strangers in the City to Tell You Their Stories. In J. M. Oakes and J. S. Kaufman, editors, Methods in Social Epidemiology, pages 239–253. Wiley: Jossey-Bass, New Jersey.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, *29*(2), 238-249.

Carrell, S. E., Hoekstra, M., & West, J. E. (2019). The impact of college diversity on behavior toward minorities. *American Economic Journal: Economic Policy*, *11*(4), 159-82.

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

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014 b). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633-79.

Cornelissen, T., Dustmann, C., & Schönberg, U. (2017). Peer effects in the workplace. *American Economic Review*, 107(2), 425-456.

Correia, S. (2016). A feasible estimator for linear models with multi-way fixed effects. *Preprint at http://scorreia. com/research/hdfe. pdf*.

Correia, S., Guimarães, P., & Zylkin, T. (2020). Fast Poisson estimation with high-dimensional fixed effects. *The Stata Journal*, *20*(1), 95-115.

Currie, Janet, and Duncan Thomas. (2001). Early test scores, school quality, and SES: Longrun effects on wage and employment outcomes. Research in Labor Economics: Worker Wellbeing in a Changing Labor Market 20:103–132. doi:10.1016/S0147-9121(01)20039-9.

Dee, Thomas S. 2004. Teachers, race, and student achievement in a randomized experiment. *The Review of Economics and Statistics* 86 (1):195-210.

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

Delgado, William. (2023). Disparate Teacher Effects, Comparative Advantage, and Match Quality. (EdWorkingPaper: 23-848). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/e81h-sp13

DeLuca, S., Clampet-Lundquist, S., & Edin, K. (2016). *Coming of age in the other America*. Russell Sage Foundation.

Foster, A. D., & Rosenzweig, M. R. (1995). Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy*, *103*(6), 1176-1209.

Gershenson, S. (2016). Linking teacher quality, student attendance, and student achievement. *Education Finance and Policy*, 11(2), 125-149.

Gershenson, S., & Langbein, L. (2015). The effect of primary school size on academic achievement. *Educational Evaluation and Policy Analysis*, *37*(1_suppl), 135S-155S.

Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student–teacher demographic match on teacher expectations. *Economics of Education Review*, *52*, 209-224.

Gershenson, S., Hansen, M. J., & Lindsay, C. A. (2021). *Teacher diversity and student success: Why racial representation matters in the classroom*. Cambridge, MA: Harvard Education Press.

Gershenson, S., Hart, C. M., Hyman, J., Lindsay, C. A., & Papageorge, N. W. (2022). The long-run impacts of same-race teachers. *American Economic Journal: Economic Policy*, *14*(4), 300-342.

Goldhaber, D., Krieg, J., & Theobald, R. (2020). Effective like me? Does having a more productive mentor improve the productivity of mentees?. *Labour Economics*, *63*, 101792.

Hamilton, B. H., Nickerson, J. A., & Owan, H. (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy*, *111*(3), 465-497.

Hamilton, B. H., Nickerson, J. A., & Owan, H. (2012). Diversity and productivity in production teams. In *Advances in the Economic Analysis of Participatory and Labor-Managed Firms* (Vol. 13, pp. 99-138). Emerald Group Publishing Limited.

Hayes, M. S., Liu, J., & Gershenson, S. (2023). Who refers whom? The effects of teacher characteristics on disciplinary office referrals. *Economics of Education Review*, *93*, 102376.

Herkenhoff, K., Lise, J., Menzio, G., & Phillips, G. M. (2024). Production and learning in teams. *Econometrica*, *92*(2), 467-504.

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

Jackson, C Kirabo. 2019. What do test scores miss? The importance of teacher effects on non-test score outcomes. *Journal of Political Economy*.

Kraft, M. A., & Papay, J. P. (2014). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis*, *36*(4), 476-500.

Kunesh, C. E., & Noltemeyer, A. (2019). Understanding disciplinary disproportionality: Stereotypes shape pre-service teachers' beliefs about black boys' behavior. *Urban Education*, *54*(4), 471-498.

Ladd, H. F., & Sorensen, L. C. (2017). Returns to teacher experience: Student achievement and motivation in middle school. *Education Finance and Policy*, *12*(2), 241-279.

Ladson-Billings, G. (2008). Yes, but how do we do it?": Practicing culturally relevant pedagogy. *City kids, city schools: More reports from the front row*, 162-177.

Ladson-Billings, G. (2022). *The Dreamkeepers: Successful teachers of African American children*. John Wiley & Sons.

Lindsay, C. A., & Hart, C. M. (2017). Exposure to same-race teachers and student disciplinary outcomes for Black students in North Carolina. *Educational Evaluation and Policy Analysis*, *39*(3), 485-510.

Lindsay, C. A., Wilson, S., Kumar, J., Byers, T., & Gershenson, S. (2024). How Teachers Learn Racial Competency: The Role of Peers and Contexts. EdWorkingPaper No. 24-968. *Annenberg Institute for School Reform at Brown University*.

Liu, J., & Loeb, S. (2021). Engaging teachers: Measuring the impact of teachers on student attendance in secondary school. *Journal of Human Resources*, *56*(2), 343-379.

Master, Benjamin, Susanna Loeb, Camille Whitney, and James Wyckoff. 2016. Different skills? Identifying differentially effective teachers of English language learners. *The Elementary School Journal*, 117 (2):261-284.

Maturana, G., & Nickerson, J. (2019). Teachers teaching teachers: The role of workplace peer effects in financial decisions. *The Review of Financial Studies*, *32*(10), 3920-3957.

Murnane, R. J., & Phillips, B. R. (1981). Learning by doing, vintage, and selection: Three pieces of the puzzle relating teaching experience and teaching performance. *Economics of Education Review*, 1(4), 453-465.

Muschkin, C., Bonneau, K., & Dodge, K. (2011). North Carolina education research data center: Final report to the Spencer Foundation (Grant# 200300138). Retrieved from http://childandfamilypolicy.duke.edu/ pdfs/projects/NCERDC_SpencerFoundationReport. pdf

Ost, Ben. (2014). How do teachers improve? The relative importance of specific and general human capital. *American Economic Journal: Applied Economics, 6*(2), 127-151.

Papageorge, N. W., Gershenson, S., & Kang, K. M. (2020). Teacher expectations matter. *Review of Economics and Statistics*, *102*(2), 234-251.

Papay, J.P., & Kraft, M.A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*. 130:105-119.

Papay, J. P., Taylor, E. S., Tyler, J. H., & Laski, M. E. (2020). Learning job skills from colleagues at work: Evidence from a field experiment using teacher performance data. *American Economic Journal: Economic Policy*, *12*(1), 359-88.

Liu, J., & Loeb, S. (2021). Engaging teachers: Measuring the impact of teachers on student attendance in secondary school. *Journal of Human Resources*, *56*(2), 343-379.

Milner, H. R., & Ford, D. Y. (2007). Cultural considerations in the underrepresentation of culturally diverse elementary students in gifted education. *Roeper Review*, *29*(3), 166-173.

Rao, G. (2019). Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools. *American Economic Review*, *109*(3), 774-809.

Sun, M., Loeb, S., & Grissom, J. A. (2017). Building teacher teams: Evidence of positive spillovers from more effective colleagues. *Educational Evaluation and Policy Analysis*, *39*(1), 104-125.

Wiswall, Matthew. 2013. The dynamics of teacher quality. *Journal of Public Economics,*100: 61-78.

Young, A. (1993). Invention and bounded learning by doing. *Journal of Political Economy*, *101*(3), 443-472.

Tables

TABLE 1. Student-year Level Summary Statistics

Sample:	All	WS WT	BS BT	BS WT
	(1)	(2)	(3)	(4)
Standardized Math Score	0.08	0.35	-0.48	-0.41
	(0.98)	(0.93)	(0.87)	(0.88)
Standardized ELA Score	0.06	0.33	-0.43	-0.37
	(0.97)	(0.91)	(0.91)	(0.91)
Absences	5.21	5.49	4.72	4.66
	(6.82)	(6.97)	(6.92)	(6.55)
Chronic Absence	0.06	0.06	0.05	0.05
Out-of-School Suspension (OSS) Days	0.12	0.05	0.29	0.25
	(1.06)	(0.71)	(1.65)	(1.44)
Ever Suspended	0.04	0.02	0.08	0.07
White	0.56	1.00	-	-
Black	0.25	-	1.00	1.00
Hispanic	0.11	-	-	-
Same Race Teacher	0.58	1.00	1.00	0.00
Female	0.48	0.48	0.49	0.50
Same Sex Teacher	0.48	0.48	0.50	0.50
N	1,669,657	863,118	101,042	307,986

Notes: W = White, B = Black, T = teachers, S = students.

TABLE 2. Teacher-year Level Summary Statistics

Variable	All	WS WT	BS BT	BS WT
	(1)	(2)	(3)	(4)
Any Black Peer	0.30	0.18	0.64	0.40
Any Black or Hispanic Peer	0.32	0.19	0.65	0.42
Class Size	17.80	18.48	16.77	17.84
	(6.70)	(6.65)	(6.45)	(6.51)
Share Black Students	0.27	0.14	0.66	0.44
	(0.26)	(0.16)	(0.24)	(0.25)
Experience 0 years	0.07	0.06	0.06	0.09
Experience | Experience>0	11.26	11.64	12.32	10.57
	(8.05)	(8.00)	(8.37)	(8.10)
Same Grade Experience 0 years	0.26	0.23	0.27	0.29
Same Grade Experience | Same Grade Experience>0	4.20	4.35	4.12	3.99
	(3.29)	(3.32)	(3.12)	(3.21)
Yrs Similar 0 years	0.30	0.27	0.30	0.37
Yrs Similar | Yrs Similar>0	3.81	3.95	3.67	3.44
	(3.03)	(3.05)	(2.89)	(2.81)
Advanced Degree	0.31	0.31	0.34	0.29
Regular License	0.84	0.87	0.84	0.81
Certified	0.11	0.13	0.05	0.10
License Exam Score	0.06	0.17	-0.61	0.15
	(0.72)	(0.67)	(0.72)	(0.67)
Female	0.90	0.91	0.91	0.89
Math Value-added	0.03	0.05	-0.03	0.03
	(0.48)	(0.51)	(0.51)	(0.48)
ELA Value-added	0.02	0.02	0.01	0.02
	(0.45)	(0.46)	(0.47)	(0.43)
Peer Experience 0 years	0.07	0.06	0.09	0.08
Peer Experience | Peer Experience>0	11.17	11.64	10.91	10.92
	(6.25)	(6.18)	(6.91)	(6.34)
Peer Advanced Degree	0.32	0.32	0.31	0.31
Peer Regular License	0.84	0.86	0.79	0.82
Peer Certified	0.11	0.12	0.06	0.09
Peer License Exam Score	0.06	0.12	-0.18	0.02
	(0.49)	(0.47)	(0.56)	(0.50)
Peer Math Value-added	0.03	0.04	0.00	0.02
	(0.34)	(0.36)	(0.36)	(0.34)
Peer ELA Value-added	0.02	0.02	0.02	0.02
	(0.31)	(0.31)	(0.32)	(0.30)
N	107,102	49,071	7,710	23,908

Notes: W = White, B = Black, T = teachers, S = students.

Table 3. Effects of Teaching Experience on EOG Math Scores

Sample:	All		Black Students of White Teachers			
	(1)	(2)	(3)	(4)	(5)	(6)
1 year	0.040***	0.031***	0.035***	0.025***	0.031***	0.024***
	(0.004)	(0.004)	(0.007)	(0.008)	(0.008)	(0.008)
2 years	0.059***	0.048***	0.065***	0.053***	0.058***	0.050***
	(0.005)	(0.005)	(0.010)	(0.010)	(0.010)	(0.010)
3 years	0.066***	0.053***	0.061***	0.047***	0.053***	0.043***
	(0.005)	(0.005)	(0.011)	(0.012)	(0.012)	(0.012)
4+ years	0.078***	0.061***	0.079***	0.059***	0.069***	0.054***
	(0.006)	(0.006)	(0.012)	(0.013)	(0.012)	(0.013)
1 yr same grade		0.015***		0.018***		0.016***
		(0.002)		(0.004)		(0.004)
2 yrs same grade		0.023***		0.022***		0.018***
		(0.002)		(0.005)		(0.006)
3 yrs same grade		0.025***		0.027***		0.024***
		(0.003)		(0.006)		(0.007)
4+ yrs same grade		0.029***		0.042***		0.037***
		(0.003)		(0.007)		(0.007)
1 yr similar					0.010***	0.007*
					(0.004)	(0.004)
2 yrs similar					0.022***	0.019***
					(0.005)	(0.005)
3 yrs similar					0.014***	0.010*
					(0.005)	(0.005)
4+ yrs similar					0.024***	0.020***
					(0.005)	(0.005)
N	1,675,005		302,474			
E(y|exp = 0)	-0.154		-0.572			

Notes: The outcome is standardized End-of-Grade (EOG) math scores. Samples contain self-contained 4th and 5th grade classrooms and are restricted to teachers entering North Carolina Public Schools in 2001 or later. The omitted reference group for all experience types is 0 years (i.e., a new teacher). Yrs similar refers to the number of years the teacher has previously taught in a classroom that has at least the same decile share of Black students in the current classroom. Regressions control for student demographics, lagged math and reading scores, class size, teacher characteristics, and fixed effects at the grade-year, school-year, and teacher-school levels. Standard errors are clustered by teacher-year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 4. Effects of White Teachers' Experience on Black Students' Educational Outcomes

Outcome:	Math EOG	Read EOG	Absences	Chronic	Days OSS	Ever OSS
	(1)	(2)	(3)	(4)	(5)	(6)
1 year	0.024***	0.008	-0.042	-0.005	-0.058***	-0.001
	(0.008)	(0.008)	(0.087)	(0.003)	(0.021)	(0.004)
2 years	0.050***	0.011	-0.242**	-0.012***	-0.050*	-0.006
	(0.010)	(0.011)	(0.117)	(0.004)	(0.029)	(0.005)
3 years	0.043***	0.030**	-0.266**	-0.014***	-0.043	-0.005
	(0.012)	(0.013)	(0.132)	(0.005)	(0.032)	(0.006)
4+ years	0.054***	0.042***	-0.444***	-0.015***	-0.107***	-0.018***
	(0.013)	(0.014)	(0.141)	(0.005)	(0.034)	(0.006)
1 yr same grade	0.016***	0.004	0.069	0.002	-0.003	-0.003
	(0.004)	(0.005)	(0.048)	(0.002)	(0.011)	(0.002)
2 yrs same grade	0.018***	-0.006	0.058	0.003	0.018	0.003
	(0.006)	(0.006)	(0.060)	(0.002)	(0.015)	(0.003)
3 yrs same grade	0.024***	-0.007	0.217***	0.008***	0.008	0.003
	(0.007)	(0.007)	(0.071)	(0.003)	(0.016)	(0.003)
4+ yrs same grade	0.037***	-0.004	0.164**	0.007**	0.031*	0.005
	(0.007)	(0.007)	(0.077)	(0.003)	(0.017)	(0.003)
1 yr similar	0.007*	0.002	-0.000	0.001	-0.021**	-0.002
	(0.004)	(0.004)	(0.042)	(0.002)	(0.010)	(0.002)
2 yrs similar	0.019***	0.016***	-0.060	-0.002	-0.042***	-0.004*
	(0.005)	(0.005)	(0.051)	(0.002)	(0.012)	(0.002)
3 yrs similar	0.010*	0.012**	0.053	-0.001	0.005	0.002
	(0.005)	(0.006)	(0.059)	(0.002)	(0.015)	(0.003)
4+ yrs similar	0.020***	0.013**	-0.031	-0.000	-0.031**	-0.003
	(0.005)	(0.005)	(0.054)	(0.002)	(0.013)	(0.003)
N	302,474	301,937	302,738	302,738	288,404	288,404
E(y|exp = 0)	-0.57	-0.52	4.90	0.06	0.34	0.09

Notes: EOG refers to end-of-grade standardized tests. OSS refers to out-of-school suspensions. Samples contain Black students in self-contained 4th and 5th grade classrooms headed by White teachers. Samples are restricted to teachers entering North Carolina Public Schools in 2001 or later. The omitted reference group for all experience types is 0 years (i.e., a new teacher). Yrs similar refers to the number of years the teacher has previously taught in a classroom that has at least the same decile share of Black students in the current classroom. Regressions control for student demographics, lagged math and reading scores, class size, teacher characteristics, and fixed effects at the grade-year, school-year, and teacher-school levels. Standard errors are clustered by teacher-year.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 5. Effects of Teaching Experience on EOG Math Scores for Different Subgroups

Sample:	All	White S White T	Black S Black T	Black S White T
	(1)	(2)	(3)	(4)
1 year	0.031***	0.034***	0.006	0.024***
	(0.004)	(0.005)	(0.018)	(0.008)
2 years	0.047***	0.049***	-0.009	0.050***
	(0.005)	(0.007)	(0.024)	(0.010)
3 years	0.051***	0.057***	-0.043*	0.043***
	(0.006)	(0.008)	(0.026)	(0.012)
4+ years	0.059***	0.066***	-0.027	0.054***
	(0.006)	(0.008)	(0.028)	(0.013)
1 yr same grade	0.015***	0.015***	0.001	0.016***
	(0.002)	(0.003)	(0.009)	(0.004)
2 yrs same grade	0.021***	0.022***	0.010	0.018***
	(0.003)	(0.003)	(0.012)	(0.006)
3 yrs same grade	0.023***	0.025***	0.010	0.024***
	(0.003)	(0.004)	(0.014)	(0.007)
4+ yrs same grade	0.026***	0.027***	0.005	0.037***
	(0.003)	(0.004)	(0.015)	(0.007)
1 yr similar	0.002	0.003	0.020**	0.007*
	(0.002)	(0.003)	(0.009)	(0.004)
2 yrs similar	0.008***	0.010***	0.022**	0.019***
	(0.002)	(0.003)	(0.010)	(0.005)
3 yrs similar	0.006**	0.007**	0.031***	0.010*
	(0.002)	(0.003)	(0.012)	(0.005)
4+ yrs similar	0.012***	0.011***	0.030***	0.020***
	(0.002)	(0.003)	(0.011)	(0.005)
N	1,675,005	856,637	104,573	302,474
E(y|exp = 0)	-0.154	0.160	-0.623	-0.572

Notes: The outcome is standardized End-of-Grade (EOG) math scores. Samples contain self-contained 4th and 5th grade classrooms and are restricted to teachers entering North Carolina Public Schools in 2001 or later for different combinations of student (S) and teacher (T) race. The omitted reference group for all experience types is 0 years (i.e., a new teacher). Yrs similar refers to the number of years the teacher has previously taught in a classroom that has at least the same decile share of Black students in the current classroom. Regressions control for student demographics, lagged math and reading scores, class size, teacher characteristics, and fixed effects at the grade-year, school-year, and teacher-school levels. Standard errors are clustered by teacher-year.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 6. Teacher Peer Effects on EOG Math Scores

Teacher Experience	All	All	Novice	Veteran
	(1)	(2)	(3)	(4)
A. All Students & All Teachers				
1{Black peer}		0.004	0.024**	0.002
		(0.003)	(0.011)	(0.003)
Peer math VAM	0.042***	0.042***	0.064***	0.041***
	(0.003)	(0.003)	(0.024)	(0.004)
N	1,731,366		318,374	1,412,779
E(y|no Black peer)	0.137		0.004	0.166
B. White Teachers' Black Students				
1{Black peer}		0.016***	0.057***	0.007
		(0.006)	(0.022)	(0.006)
Peer math VAM	0.048***	0.049***	0.004	0.048***
	(0.007)	(0.007)	(0.041)	(0.008)
N	314,657		70,735	242,743
E(y|no Black peer)	-0.390		-0.490	-0.363
C. Peer Experience, for White Teachers' Black Students				
Veteran Black Peer		0.015**	0.077***	0.006
		(0.006)	(0.024)	(0.007)
Novice Black Peer		0.020**	0.002	0.012
		(0.009)	(0.036)	(0.011)
Peer math VAM		0.049***	0.001	0.048***
		(0.007)	(0.041)	(0.008)
D. Peer Effectiveness, for White Teachers' Black Students				
More Effective Black Peer		0.043***	0.135**	0.034*
		(0.016)	(0.067)	(0.019)
Less Effective Black Peer		0.015**	0.053**	0.006
		(0.006)	(0.023)	(0.007)
Peer math VAM		0.046***	-0.014	0.046***
		(0.007)	(0.040)	(0.008)

Notes: The outcome is standardized End-of-Grade (EOG) math scores. Samples contain self-contained 4th and 5th grade classrooms in North Carolina Public Schools from 2001 to 2018. Regressions control for student demographics, lagged math and reading scores, class size, teacher characteristics, teachers' *average* peer characteristics, and fixed effects at the grade-year, school-year, and teacher-school levels. Peers are defined at the school-grade-year level. 1{Black Peer} is a binary indicator equal to one if a teacher had a Black peer in year *t*, and 0 otherwise. Novice is defined as 2 or fewer years. Average peer value-added measures (VAM) are computed using pre-2001 data. Standard errors are clustered by teacher-year. In panel D More effective is defined as > 1 SD above the mean. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 7. White Teachers' Peers' Effects on Black Students' Educational Outcomes

	All	Novice	Veteran	All	Novice	Veteran
	(1)	(2)	(3)	(4)	(5)	(6)
Outcome		EOG Math			EOG Reading	
1{Black peer}	0.016***	0.057***	0.007	0.011*	0.045*	0.005
	(0.006)	(0.022)	(0.006)	(0.006)	(0.024)	(0.007)
N	314,657	70,735	242,743	313,983	70,592	242,211
E(y|no Black peer)	-0.39	-0.47	-0.36	-0.35	-0.43	-0.32
		Annual Absences			1{Chronically Absent}	
1{Black peer}	-0.089	-0.624***	-0.050	-0.005**	-0.023***	-0.003
	(0.059)	(0.209)	(0.069)	(0.002)	(0.008)	(0.003)
N	314,917	70,798	242,939	314,917	70,798	242,939
E(y|no Black peer)	4.7	4.8	4.6	0.05	0.05	0.05
		OSS Days			1{Ever OSS}	
1{Black peer}	0.010	-0.112**	0.004	-0.005*	-0.019*	-0.006*
	(0.014)	(0.053)	(0.016)	(0.003)	(0.011)	(0.003)
N	301,782	69,960	230,689	301,782	69,960	230,689
E(y|no Black peer)	0.24	0.29	0.23	0.07	0.08	0.07

Notes: EOG refers to end-of-grade standardized tests. OSS refers to out-of-school suspensions Samples contain self-contained 4th and 5th grade classrooms in North Carolina Public Schools from 2001 to 2018. Regressions control for student demographics, lagged math and reading scores, class size, teacher characteristics, teachers' *average* peer characteristics (including *both* math and reading VAM), and fixed effects at the grade-year, school-year, and teacher-school levels. Peers are defined at the school-grade-year level. 1{Black Peer} is a binary indicator equal to one if a teacher had a Black peer in year *t*, and 0 otherwise. Novice is defined as 2 or fewer years. Standard errors are clustered by teacher-year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 8. Teacher Peer Effects on EOG Math Scores for Different Subgroups

	Black students of Black teachers			White students of White teachers		
	All	Novice	Veteran	All	Novice	Veteran
	(1)	(2)	(3)	(4)	(5)	(6)
1{Black peer}	-0.026**	-0.183	-0.020*	-0.003	0.001	-0.004
	(0.011)	(0.127)	(0.012)	(0.004)	(0.020)	(0.005)
Peer math VAM	-0.001	0.016	-0.007	0.044***	0.060*	0.043***
	(0.014)	(0.201)	(0.016)	(0.004)	(0.035)	(0.004)
N	107,893	16,286	91,560	889,771	142,395	746,812
E(y|no Black peer)	-0.47	-0.53	-0.46	0.34	0.23	0.36

Notes: The outcome is standardized End-of-Grade (EOG) math scores. Samples contain self-contained 4th and 5th grade classrooms in North Carolina Public Schools from 2001 to 2018. Regressions control for student demographics, lagged math and reading scores, class size, teacher characteristics, teachers' *average* peer characteristics, and fixed effects at the grade-year, school-year, and teacher-school levels. Peers are defined at the school-grade-year level. 1{Black Peer} is a binary indicator equal to one if a teacher had a Black peer in year *t*, and 0 otherwise. Novice is defined as 2 or fewer years. Average peer value-added measures (VAM) are computed using pre-2001 data. Standard errors are clustered by teacher-year.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 9. Leads and Lags of Teacher Peer Effects on EOG Math Scores

Sample:	1 Lag	2 Lags	1 Lead	2 Leads	Lead and Lag
	(1)	(2)	(3)	(4)	(5)
A. Baseline Model					
1{Black peer}	0.019**	0.005	0.016*	0.011	0.010
	(0.008)	(0.010)	(0.008)	(0.012)	(0.011)
Peer math VAM	0.046***	0.041***	0.044***	0.035***	0.036***
	(0.009)	(0.012)	(0.010)	(0.013)	(0.012)
B. Leads and Lags					
1{Black peer}	0.019**	0.005	0.017**	0.010	0.010
	(0.008)	(0.010)	(0.008)	(0.012)	(0.011)
Lag 1{Black peer}	-0.004	0.005			-0.002
	(0.007)	(0.010)			(0.010)
Second Lag 1{Black peer}		0.024**			
		(0.009)			
Lead 1{Black peer}			-0.006	0.007	0.001
			(0.008)	(0.012)	(0.011)
Second Lead 1{Black peer}				-0.009	
				(0.011)	
Peer math VAM	0.040***	0.036***	0.043***	0.036***	0.030**
	(0.009)	(0.012)	(0.010)	(0.013)	(0.012)
Lag Peer math VAM	0.026***	0.018			0.023**
	(0.009)	(0.011)			(0.011)
Second Lag Peer math VAM		0.031***			
		(0.010)			
Lead Peer math VAM			0.008	-0.009	0.007
			(0.010)	(0.013)	(0.012)
Second Lead Peer math VAM				-0.002	
				(0.014)	
Observations	198,716	131,493	170,607	101,066	116,797
E(y|no Black peer)	-0.35	-0.33	-0.35	-0.34	-0.33

Notes: The outcome is standardized End-of-Grade (EOG) math scores. Samples contain self-contained 4th and 5th grade classrooms in North Carolina Public Schools from 2001 to 2018. Sample sizes change because the number of leads and/or lags change the data requirements; Panel A estimates the baseline model (no leads or lags) on the same restricted analytic sample. Regressions control for student demographics, lagged math and reading scores, class size, teacher characteristics, teachers' *average* peer characteristics, and fixed effects at the grade-year, school-year, and teacher-school levels. Peers are defined at the school-grade-year level. 1{Black Peer} is a binary indicator equal to one if a teacher had a Black peer in year *t*, and 0 otherwise. Average peer value-added measures (VAM) are computed using pre-2001 data. Standard errors are clustered by teacher-year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Online Appendix

Appendix Table A1. Effects of Teaching Experience on EOG Reading Scores

Sample:	All		Black Students of White Teachers			
	(1)	(2)	(3)	(4)	(5)	(6)
1 year	0.015***	0.011***	0.012	0.009	0.011	0.008
	(0.003)	(0.003)	(0.008)	(0.008)	(0.008)	(0.008)
2 years	0.025***	0.020***	0.013	0.013	0.010	0.011
	(0.004)	(0.005)	(0.011)	(0.011)	(0.011)	(0.011)
3 years	0.029***	0.024***	0.033***	0.034***	0.029**	0.030**
	(0.005)	(0.005)	(0.012)	(0.012)	(0.012)	(0.013)
4+ years	0.037***	0.031***	0.047***	0.047***	0.041***	0.042***
	(0.005)	(0.005)	(0.013)	(0.014)	(0.013)	(0.014)
1 yr same grade		0.006***		0.005		0.004
		(0.002)		(0.004)		(0.005)
2 yrs same grade		0.008***		-0.003		-0.006
		(0.002)		(0.006)		(0.006)
3 yrs same grade		0.009***		-0.003		-0.007
		(0.003)		(0.007)		(0.007)
4+ yrs same grade		0.008***		-0.001		-0.004
		(0.003)		(0.007)		(0.007)
1 yr similar					0.003	0.002
					(0.004)	(0.004)
2 yrs similar					0.015***	0.016***
					(0.005)	(0.005)
3 yrs similar					0.011*	0.012**
					(0.006)	(0.006)
4+ yrs similar					0.013**	0.013**
					(0.005)	(0.005)
N	1,672,468		301,937			
E(y|exp = 0)	-0.146		-0.520			

Notes: The outcome is standardized End-of-Grade (EOG) math scores. Samples contain self-contained 4th and 5th grade classrooms and are restricted to teachers entering North Carolina Public Schools in 2001 or later. The omitted reference group for all experience types is 0 years (i.e., a new teacher). Yrs similar refers to the number of years the teacher has previously taught in a classroom that has at least the same decile share of Black students in the current classroom. Regressions control for student demographics, lagged math and reading scores, class size, teacher characteristics, and fixed effects at the grade-year, school-year, and teacher-school levels. Standard errors are clustered by teacher-year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Appendix Table A2. Effects of Experience Sensitivity to how Standard Errors are Clustered

Sample:	All		Black Students of White Teachers			
	(1)	(2)	(3)	(4)	(5)	(6)
Experience	-0.002	-0.002	-0.001	-0.002	-0.002	-0.002
	(0.002)	(0.002)	(0.004)	(0.004)	(0.004)	(0.004)
1{New to NC}	-0.043***	-0.031***	-0.038***	-0.027***	-0.032***	-0.025***
Teacher-year	(0.004)	(0.004)	(0.007)	(0.008)	(0.008)	(0.008)
Teacher	(0.004)	(0.004)	(0.009)	(0.009)	(0.009)	(0.009)
School-year	(0.004)	(0.004)	(0.008)	(0.008)	(0.008)	(0.009)
School	(0.005)	(0.005)	(0.010)	(0.010)	(0.010)	(0.010)
T-year & Student	(0.004)	(0.004)	(0.007)	(0.008)	(0.008)	(0.008)
T & S	(0.004)	(0.004)	(0.009)	(0.009)	(0.009)	(0.009)
Sch-yr & Student	(0.004)	(0.004)	(0.008)	(0.008)	(0.008)	(0.009)
School & Student	(0.005)	(0.005)	(0.010)	(0.010)	(0.010)	(0.010)
Same grade Exp.		0.001		0.004***		0.004***
		(0.001)		(0.001)		(0.001)
1{New to grade}		-0.018***		-0.016***		-0.013***
Teacher-year		(0.002)		(0.004)		(0.004)
Teacher		(0.002)		(0.005)		(0.005)
School-year		(0.002)		(0.005)		(0.005)
School		(0.003)		(0.006)		(0.006)
T-year & Student		(0.002)		(0.004)		(0.004)
T & S		(0.002)		(0.005)		(0.005)
Sch-yr & Student		(0.002)		(0.005)		(0.005)
School & Student		(0.003)		(0.006)		(0.006)
Similar Class Exp					0.001	0.001
					(0.001)	(0.001)
1{New to class}					-0.014***	-0.011***
Teacher-year					(0.004)	(0.004)
Teacher					(0.004)	(0.004)
School-year					(0.004)	(0.004)
School					(0.005)	(0.005)
T-year & Student					(0.004)	(0.004)
T & S					(0.004)	(0.004)
Sch-yr & Student					(0.004)	(0.004)
School & Student					(0.005)	(0.005)
N	1,675,005		302,474			
E(y|exp = 0)	-0.154		-0.572			

Notes: This table is analogous to Table 3, but replace the nonparametric experience indicators with continuous measures and an indicator for having zero years of a particular type of experience. The outcome is standardized End-of-Grade (EOG) math scores. Samples contain self-contained 4th and 5th grade classrooms and are restricted to teachers entering North Carolina Public Schools in 2001 or later. Yrs similar refers to the number of years the teacher has previously taught in a classroom that has at least the same decile share of Black students in the current classroom. Regressions control for student demographics, lagged math and reading scores, class size, teacher characteristics, and fixed effects at the grade-year, school-year, and teacher-school levels. Standard errors are clustered by teacher-year, except where otherwise noted. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Appendix Table A3. Poisson Regressions for Count Outcomes

Teachers	All	Novice	Veteran	All	Novice	Veteran
Outcome		Annual Absences			OSS Days	
	(1)	(2)	(3)	(4)	(5)	(6)
A. Experience						
1 year	-0.023			-0.143*		
	(0.015)			(0.085)		
2 years	-0.066***			-0.133		
	(0.021)			(0.119)		
3 years	-0.070***			-0.109		
	(0.024)			(0.131)		
4+ years	-0.102***			-0.258*		
	(0.026)			(0.149)		
1 yr same grade	0.021**			-0.043		
	(0.009)			(0.052)		
2 yrs same grade	0.018			0.055		
	(0.011)			(0.063)		
3 yrs same grade	0.055***			-0.020		
	(0.013)			(0.076)		
4+ yrs same grade	0.040***			0.033		
	(0.014)			(0.084)		
1 yr similar	-0.006			-0.138***		
	(0.008)			(0.045)		
2 yrs similar	-0.021**			-0.217***		
	(0.010)			(0.053)		
3 yrs similar	-0.003			-0.005		
	(0.012)			(0.062)		
4+ yrs similar	-0.022**			-0.199***		
	(0.011)			(0.061)		
N	234,459			128,200		
E(y|exp = 0)	6.9			0.80		
B. Peer Effects						
1{Black peer}	-0.022*	-0.169***	-0.017	-0.023	-0.687**	-0.045
	(0.012)	(0.043)	(0.014)	(0.073)	(0.285)	(0.091)
N	225,649	50,768	173,850	124,856	25,385	91,104
E(y|no Black peer)	6.4	6.7	6.3	0.63	0.85	0.62

Notes: All samples are restricted to the Black students of white teachers. In Panel A samples are further restricted to teachers entering North Carolina Public Schools in 2001 or later and the omitted reference group for all experience types is 0 years (i.e., a new teacher). Yrs similar refers to the number of years the teacher has previously taught in a classroom that has at least the same decile share of Black students in the current classroom. 1{Black Peer} is a binary indicator equal to one if a teacher had a Black peer in year *t*, and 0 otherwise. Regressions control for student demographics, lagged math and reading scores, class size, teacher characteristics, teachers' peers' average characteristics, and fixed effects at the grade-year, school-year, and teacher-school levels. Standard errors are clustered by teacher-year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Appendix Table A4. Teacher Peer Effects on EOG Reading Scores

Teacher Experience	All	All	Novice	Veteran
	(1)	(2)	(3)	(4)
A. All Students				
1{Black peer}		0.003	0.014	0.002
		(0.002)	(0.010)	(0.003)
Peer read VAM	0.023***	0.023***	0.012	0.024***
	(0.003)	(0.003)	(0.018)	(0.003)
N	1,728,075		317,813	1,410,046
E(y|no Black peer)	0.116		-0.008	0.143
B. White Teachers' Black Students				
1{Black peer}		0.011*	0.044*	0.004
		(0.006)	(0.024)	(0.007)
Peer read VAM	0.032***	0.033***	0.014	0.027***
	(0.007)	(0.008)	(0.046)	(0.008)
N	313,983		70,592	242,211
E(y|no Black peer)	-0.353		-0.448	-0.327

Notes: The outcome is standardized End-of-Grade (EOG) reading scores. Samples contain self-contained 4th and 5th grade classrooms in North Carolina Public Schools from 2001 to 2018. Regressions control for student demographics, lagged math and reading scores, class size, teacher characteristics, teachers' *average* peer characteristics, and fixed effects at the grade-year, school-year, and teacher-school levels. Peers are defined at the school-grade-year level. 1{Black Peer} is a binary indicator equal to one if a teacher had a Black peer in year *t*, and 0 otherwise. Novice is defined as 2 or fewer years. Average peer value-added measures (VAM) are computed using pre-2001 data. Standard errors are clustered by teacher-year.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Appendix Table A5. Novice Teacher Peer Effects on EOG Reading Scores

Novice Definition	< 1 yr	< 2 yrs	< 3 yrs	< 4 yrs	< 5 yrs	< 6 yrs
	(1)	(2)	(3)	(4)	(5)	(6)
A. All Students and Teachers						
1{Black peer}	0.033	0.019	0.024**	0.015*	0.014**	0.011*
	(0.052)	(0.017)	(0.011)	(0.008)	(0.007)	(0.006)
Peer math VAM	0.118	0.054	0.064***	0.057***	0.064***	0.058***
	(0.090)	(0.037)	(0.024)	(0.017)	(0.012)	(0.010)
N	116,567	219,219	318,374	414,883	505,559	590,419
E(y|no Black peer)	-0.07	-0.03	0.004	0.03	0.05	0.06
B. Black Students of White Teachers						
1{Black peer}	0.026	0.033	0.057***	0.039**	0.047***	0.049***
	(0.111)	(0.032)	(0.022)	(0.018)	(0.015)	(0.013)
Peer math VAM	-0.195	0.019	0.004	-0.006	0.021	0.031
	(0.201)	(0.105)	(0.041)	(0.030)	(0.023)	(0.019)
N	26,512	49,499	70,735	90,568	108,554	124,856
E(y|no Black peer)	-0.55	-0.52	-0.49	-0.47	-0.46	-0.44

Notes: This Table replicates column 3 of Table 6 using different definitions of "novice." The outcome is standardized End-of-Grade (EOG) math scores. Samples contain self-contained 4th and 5th grade classrooms in North Carolina Public Schools from 2001 to 2018. Regressions control for student demographics, lagged math and reading scores, class size, teacher characteristics, teachers' *average* peer characteristics, and fixed effects at the grade-year, school-year, and teacher-school levels. Peers are defined at the school-grade-year level. 1{Black Peer} is a binary indicator equal to one if a teacher had a Black peer in year *t*, and 0 otherwise. Average peer value-added measures (VAM) are computed using pre-2001 data. Standard errors are clustered by teacher-year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Appendix Table A6. Peer Effects Sensitivity to How Standard Errors Are Clustered

Teacher Experience	All	All	Novice	Veteran
	(1)	(2)	(3)	(4)
A. All Students				
1{Black peer}		0.004	0.024**	0.002
Teacher-year		(0.003)	(0.011)	(0.003)
Teacher		(0.003)	(0.014)	(0.003)
School-year		(0.003)	(0.013)	(0.004)
School		(0.004)	(0.017)	(0.005)
T-year & Student		(0.003)	(0.011)	(0.003)
T & S		(0.003)	(0.014)	(0.003)
Sch-yr & Student		(0.003)	(0.013)	(0.004)
School & Student		(0.004)	(0.017)	(0.005)
Peer math VAM	0.042***	0.042***	0.064***	0.041***
Teacher-year	(0.003)	(0.003)	(0.024)	(0.004)
Teacher	(0.004)	(0.004)	(0.027)	(0.004)
School-year	(0.004)	(0.004)	(0.027)	(0.004)
School	(0.005)	(0.005)	(0.030)	(0.006)
T-year & Student	(0.003)	(0.003)	(0.024)	(0.004)
T & S	(0.004)	(0.004)	(0.027)	(0.004)
Sch-yr & Student	(0.004)	(0.004)	(0.027)	(0.005)
School & Student	(0.005)	(0.005)	(0.030)	(0.006)
B. White Teachers' Black Students				
1{Black peer}		0.016***	0.057***	0.007
Teacher-year		(0.006)	(0.022)	(0.006)
Teacher		(0.006)	(0.027)	(0.007)
School-year		(0.007)	(0.026)	(0.007)
School		(0.007)	(0.033)	(0.009)
T-year & Student		(0.006)	(0.022)	(0.007)
Teacher & Student		(0.006)	(0.027)	(0.007)
Sch-yr & Student		(0.007)	(0.026)	(0.008)
School & Student		(0.007)	(0.033)	(0.009)
Peer math VAM	0.048***	0.049***	0.004	0.048***
Teacher-year	(0.007)	(0.007)	(0.041)	(0.008)
Teacher	(0.008)	(0.008)	(0.048)	(0.009)
School-year	(0.008)	(0.008)	(0.047)	(0.009)
School	(0.010)	(0.010)	(0.055)	(0.011)
T-year & Student	(0.007)	(0.007)	(0.041)	(0.008)
T & S	(0.008)	(0.008)	(0.048)	(0.009)
Sch-yr & Student	(0.008)	(0.008)	(0.047)	(0.009)
School & Student	(0.010)	(0.010)	(0.055)	(0.011)

Note: This table is identical to Table 6, with seven additional standard errors clustered at successively higher levels (or two-way clustered by student as well).

Appendix Table A7. Teacher Peer Effects on EOG Math Scores

Teacher Experience	All	All	Novice	Veteran
	(1)	(2)	(3)	(4)
A. All Students				
% Black Peers		0.003	0.030	0.000
		(0.006)	(0.026)	(0.006)
Peer math VAM	0.042***	0.042***	0.064***	0.041***
	(0.003)	(0.003)	(0.024)	(0.004)
N	1,731,366		318,374	1,412,779
E(y|no Black peer)	0.137		0.004	0.166
B. White Teachers' Black Students				
% Black Peers		0.014	0.073	-0.000
		(0.011)	(0.053)	(0.013)
Peer math VAM	0.048***	0.048***	0.002	0.047***
	(0.007)	(0.007)	(0.041)	(0.008)
N	314,657		70,735	242,743
E(y|no Black peer)	-0.390		-0.490	-0.363

Notes: This Table replicates Table 6 but changes from an indicator for having at least one Black peer to the fraction of one's peers who are Black. The outcome is standardized End-of-Grade (EOG) math scores. Samples contain self-contained 4th and 5th grade classrooms in North Carolina Public Schools from 2001 to 2018. Regressions control for student demographics, lagged math and reading scores, class size, teacher characteristics, teachers' *average* peer characteristics, and fixed effects at the grade-year, school-year, and teacher-school levels. Peers are defined at the school-grade-year level. Novice is defined as 2 or fewer years. Average peer value-added measures (VAM) are computed using pre-2001 data. Standard errors are clustered by teacher-year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.