Skill-Specific Productivity and On-the-Job Learning: How Peers and Experience Improve White Teachers of Black Students

Romaine A. Campbell[1]
Seth Gershenson[2]
Constance A. Lindsay[3]
Nicholas W. Papageorge[4]

Abstract: Workers learn on the job from their peers and from experience, e.g., trial and error. Less understood is how specific types of peers and experiences affect on-the-job learning, which likely differs by context (e.g., occupation, tasks, or roles). Absent such knowledge, it is unclear how to optimally assign workers to specific teams and tasks. We examine on-the-job learning among elementary school teachers. We focus on the nuanced case of white teachers' productivity teaching Black students. We examine specific types of peers and classroom experiences that could lead to rapid productivity gains for white teachers in diverse classrooms: having Black colleagues and accruing experience teaching Black students. Both lead to significant productivity gains over and above those associated with access to generally productive peers or general teaching experience. This is due to learning, as peer effects are persistent and driven by more effective and experienced 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.

Keywords: Human Capital, Peer Effects, Knowledge Spillovers, Experience, Learning-by-Doing, Teacher Effectiveness, Teacher Diversity, Racial Achievement Gaps, Education Production Function.

JEL Codes: I2, J24, D2

---

[1] Department of Economics and Brooks School of Public Policy, Cornell University (rcampbell@cornell.edu).
Campbell gratefully acknowledges financial support from the National Science Foundation.
[2] Corresponding author. School of Public Affairs, American University and IZA (seth.gershenson@american.edu).
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.
[3] School of Education, University of North Carolina at Chapel Hill (clindsay@unc.edu).
[4] Department of Economics, Johns Hopkins University, IZA, and NBER (papageorge@jhu.edu).

1. Introduction

A vast literature on human capital accumulation studies on-the-job learning, i.e., how workers become more productive throughout their careers (Becker 1962; Mincer 1962). Outside of formal training, two mechanisms are dominant: learning-by-doing, e.g., through trial-and-error (Arrow 1962; Ben-Porath 1967; Levitt et al. 2013); and peer effects driven by knowledge spillovers (Cornelissen et al. 2017). Both peer effects and experience can increase productivity in general (Foster & Rosenzweig 1995; Herkenhoff et al. 2024; Jackson & Bruegmann 2009; Young 1993). However, many organizations and firms are best served not by increasing workers' general productivity, but by developing nuanced skills relevant to specific tasks, which we refer to as *skill-specific productivity*. We know far less about how workers gain *skill-specific productivity* through on-the-job learning, including the roles that peers and experiences play in this process, which we refer to as *skill-specific peers* and *skill-specific experience*. Absent such knowledge, it is difficult to efficiently allocate workers to the teams or tasks that—given a firm's or organization's goals—optimally leverage on-the-job learning opportunities.

We examine how skill-specific peers and experiences affect productivity in a particular context: white teachers' ability to teach Black students. We investigate how both Black peers and past experience with Black students affect white teachers' productivity teaching Black students. This context is important on its own: Black students benefit when they have a Black teacher, yet the teaching force is disproportionately white (Dee 2004; Gershenson et al. 2021, 2022). Achieving a more representative teaching force is a reasonable and sensible policy response, but in the short run—and absent major and potentially disruptive or unrealistic shifts to the teacher labor market—it is unlikely to appreciably change how many Black students will encounter a Black teacher. We thus investigate whether the unique skills underlying race-match effects can be acquired by white teachers. This context provides an ideal setting to examine skill-specific on-the-job learning since it is straightforward to measure skill-specific productivity, peers, and experience in schools.

Using administrative data from North Carolina, we measure productivity gains as changes in teachers' impacts on students' standardized test scores, attendance rates, and suspensions. To study the impact of skill-specific peers, we examine the impact of having Black colleagues (who are well-documented as particularly effective teachers of Black students (see, e.g., Gershenson et al. 2021, 2022)) on white teachers' performance teaching Black students. The

idea is that Black teachers possess specific skills, insights, experience, or knowledge relevant to teaching Black students (e.g., Ladson-Billings 2022) that white teachers can acquire through knowledge spillovers. To study the impact of skill-specific experience, we examine the impact of white teachers' prior experience teaching Black students on their performance teaching Black students. Here, the idea is that experience with Black students, including trial-and-error, could improve how white teachers approach teaching diverse classrooms. Finally, we augment these quantitative analyses with qualitative data collected from semi-structured interviews of teachers in North Carolina, which were collected and analyzed with the goal of better understanding the estimated coefficients. Specifically, the interview data help to corroborate our general interpretation, demystify puzzling results, and suggest mechanisms that may drive effects.

We provide three main sets of findings. First, having a same-grade Black peer increases white teachers' effectiveness teaching Black students, improving math and reading scores by 1% of a test-score standard deviation (SD). Moreover, increases are larger among novice white teachers (and their Black students) when they have a Black same-grade teacher peer: math and reading scores of Black students increase by 6% and 4% of a SD, respectively. That peer effects are stronger among novice teachers comports with prior research on teacher peer effects (Jackson & Bruegmann 2009; Maturana & Nickerson 2019) and suggests that peer learning is the primary mechanism since new teachers are honing their practice. Moreover, these effects are persistent: estimates of historical exposure to Black peers remain statistically significant and larger in magnitude than contemporaneous exposure. Furthermore, we find no effect of having a Black peer on white teachers' white students' achievement, suggesting a specific, rather than general, skill is being transmitted. Finally, falsification tests show no impact of having a future Black peer on Black students' current outcomes, bolstering a causal interpretation of our results through knowledge spillovers related to a specific skill.

Second, we show Black students perform better when assigned to white teachers with prior experience teaching Black students. Holding total teaching experience constant, the math scores of white teachers' Black students increase by an additional 1 to 3% of a SD when their teacher has at least one year of experience in a classroom that was at least 25% Black. The mechanism could be trial and error but could also work through reductions in implicit biases or increases in cultural competence (Papageorge et al. 2020; Dee & Penner 2017). Yet, in contrast to the case of Black peers, we also find that white students benefit (roughly half as much as

Black students) from white teachers' experience having taught Black students. In other words, while Black colleagues may provide white teachers with highly specific knowledge spillovers that improve their performance solely with Black students, experience teaching in a diverse classroom improves white teachers' *general* productivity too. An interpretation, supported by the qualitative data, is that these teachers become better equipped to address the novel needs of individual students, consistent with a disruption model of education production (Lazear 2001).

Third, we augment our quantitative analyses with qualitative analyses of open-ended, in-depth interviews of 39 teachers. Qualitative data are rarely used in economics, though a handful of examples suggest that they can be used to aid in model specification, explain puzzling empirical patterns, or suggest new hypotheses to test with larger data sets (e.g., Bergman et al. 2024; Bewley 1995; DeLuca et al. 2024, 2026). Our use of qualitative data addresses two simple questions: First, when prompted to discuss how they learn on the job, do teachers corroborate our empirical findings, i.e., do they mention challenges teaching students with backgrounds different from their own or the benefits of skill-specific experience and knowledgeable peers? Second, do interviews with teachers provide any additional insights about how on-the-job learning works in the context we study that would help to understand mechanisms or counterintuitive results?

According to the qualitative data, white teachers learn directly from their Black peers through informal conversations and observations as well as through formal workshops and mentorships. They also report learning through formative student interactions that enable them to learn about race and other elements of student identity. Moreover, the qualitative data offer a unique addition that speaks to mechanisms: some teachers who report learning from peers mention that (i) conversations regarding racial competency in the classroom can be delicate and difficult and that (ii) it is important peers approach these conversations with kindness, empathy, and a lack of judgment to facilitate learning. This finding is policy relevant, as it suggests not only that peer learning is important, but that training related to tone and approach can enhance the quantity and quality of peer learning, leading to a more productive and dynamic teacher work force that is equipped to teach in increasingly diverse classrooms.

Our study has implications for any occupation or workplace, including teaching, in which there are unique types of experience and peers that increase productivity for specific tasks. Workers in any organization are likely to face challenges performing certain tasks because of their own inexperience or under-developed task-specific skills, even when they are productive in

other domains and in the aggregate. Our findings provide evidence of the power of on-the-job learning that can be far more effective with exposure to skill-specific experience and skill-specific peers, both of which can accelerate improvements in skill-specific productivity.

Moreover, that white teachers can improve how well they teach Black students is a striking example of skill-specific on-the-job learning. Race match effects are well documented (Dee 2004, Gershenson et al. 2022), but the skills undergirding them are often viewed as difficult to transfer to the white teachers who compose the majority of the teacher workforce (e.g., D'Amico et al. 2017). Our findings on on-the-job learning suggest they are transferable and point to a crucial role for strategic placement of Black teachers that leverages two key positive impacts Black teachers have on Black students: direct and well-documented "race match" effects, along with indirect "knowledge spillover" effects via their interactions with white colleagues who teach Black students. Our findings also underscore the importance of experience: white teachers can overcome struggles with diverse classrooms' needs via a process of learning by doing. A possible channel is through reductions in implicit biases, which earlier literature suggests can be reduced with information and exposure (Billings et al. 2021; Carrell et al. 2019; Rao et al. 2019).

Our study contributes to three main literatures. The first is on production theory and workforce teams. A typical finding is that diverse-in-ability teams are more productive (e.g., Hamilton et al. 2003). Related literature on workplace peer effects generally finds that having more productive peers increases individuals' productivity (e.g., Cornelissen et al. 2017; Mas & Moretti 2009). 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. Interestingly, having less productive peers does not have a symmetrically negative impact, suggesting that gains outweigh losses when productive employees are partnered with less productive employees. It is theoretically ambiguous whether creating racially and ethnically diverse teams is beneficial, as higher communication costs may hinder performance. Indeed, 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). Worker productivity in such a setting is likely orthogonal to race and ethnicity, but not necessarily so in schools, where research shows that teachers vary in their ability to teach students of different backgrounds (e.g., Dee 2004; Delgado 2025; Gershenson et al. 2021, 2022). We study the impact of racially diverse peers in a setting in which racial diversity is directly related to productivity. Our findings

demonstrate the importance of measuring heterogeneity in human capital accumulation via on-the-job learning across different work contexts.

We also contribute to the literature on learning by doing, which has a long history in economics as both a driver of economic growth (Arrow 1962) and of human capital accumulation (Becker 1962; Ben-Porath 1967). Learning by doing is a type of learning on the job that is distinct from knowledge spillovers associated with peer effects, teams, and mentors.[5] Rather, both individuals and teams can gain efficiency via repetition. For example, Wright (1936) shows that airplane mechanics' productivity increased during a period that saw no investments in capital or training. Individuals and teams can also learn from the successes and failures that result from successive attempts at a solution (i.e., trial and error) (Callender 2011). The benefits of trial and error may be particularly large when the production process is complex and/or agents are operating without full information, conditions that are present in classrooms (Murnane & Phillips 1981). A large literature in the economics of education has sought to identify the *returns* to teaching experience, with relatively little focus on *why* those returns exist (e.g., Wiswall 2013). We contribute to this gap by showing that teachers learn on the job via both trial and error and knowledge spillovers from their peers in ways that are unique to the *specific types* of experiences and peers encountered.

We also contribute to literature on the role of teachers in the education production function (e.g., Chetty et al. 2014 a, b; Jackson 2018) and specifically to how teachers become more productive via on-the-job learning. Papay et al. (2020) evaluate a field experiment in which more and less productive teachers were intentionally matched in a mentorship type arrangement and find that the students of the less productive teacher improved by 12% of a SD.[6] Jackson and Bruegmann (2009) find significant and long-lasting increases in teacher effectiveness when they are exposed to effective same-grade peers (measured by value-added scores). These teacher peer effects are most apparent among novice teachers and are persistent over time, suggesting they are driven by information transmission, or knowledge spillovers. Relatedly, Sun et al. (2017) show that when an effective teacher changes schools, she increases the effectiveness of her new

[5] Of course, learning-by-doing and peer effects can occur in tandem and are not mutually exclusive avenues to increasing human capital over the course of one's career (Foster & Rosenzweig 1995; Young 1993). Indeed, in results available upon request, we also estimate whether in our context these two modes of on-the-job learning interact, i.e., whether they are complements or substitutes, but evidence is inconclusive.

[6] A similar effect is found when preservice teachers are matched to more productive mentor teachers for their student teaching experience (Goldhaber et al. 2020).

colleagues. Finally, Maturana and Nickerson (2019) find that teachers, especially younger ones, are more likely to refinance their mortgage when their colleagues do, suggesting the presence of knowledge spillovers in domains outside their primary scope of work. We contribute to this literature by showing that race-specific teaching skills (e.g., Delgado 2025; Gershenson et al. 2022) are not fixed and can be transferred to white teachers from Black peers.

Regarding gains from experience, Murnane and Phillips (1981) speculated that teachers improve over time via "learning by doing." Since then, a large literature has documented that teachers do, on average, improve significantly over the course of their careers (e.g., Wiswall 2013; Papay & Kraft 2015; Bell et al. 2025). Moreover, the returns to teaching experience are larger in more supportive environments (Kraft & Papay 2014), which could indicate complementarities with peer effects. Ost (2014) distinguishes between general and specific human capital, in the sense of Becker (1962), by identifying extra returns to grade-specific experience. Descriptively, novice teachers tend to be more productive when they teach in schools that resemble where they student taught (Goldhaber et al. 2017). 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 further evidence that skill-specific experience is valuable. Our findings are also consistent with the delineation between general and specific human capital. Black peers only affect white teachers' specific ability to teach Black students, while experience teaching Black students builds a more general form of human capital that benefits white students too, albeit to a lesser extent. Teacher allocation to classrooms and peers could leverage either or both types of on-the-job learning, depending on teachers' human capital and the school's teaching force and specific needs.

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 peer effects and the returns to experience, 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 data security requirements (Gershenson & Langbein 2015; Muschkin et al. 2011). These are the same data analyzed by Ost (2014) and 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 matched to a self-contained classroom in grades 3 through 5 between 2001 and 2018 in North Carolina public schools while column 4 does so for our primary analytic sample of 314,657 student-year observations for Black students matched to white teachers. 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' scoring lower than 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 24,074 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 6.4 and classrooms, on average, contain 45 percent Black students. About 30 percent of teachers hold an advanced degree. Most teachers in our sample (82 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 0.16 with a standard deviation of 1.27. Approximately 89 percent of teachers are female. Teachers have average value-added of 0.03 and 0.02 test-score standard deviations (again, this is standardized) with standard deviations of 0.49 and 0.44 for math and ELA, respectively.

2.1 Peer Characteristics

Approximately 41 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 41% 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 (2009), 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 about 11 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 28 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 Black students. We somewhat arbitrarily define such past experience as years in classrooms with at least 25 percent Black students, the full sample mean observed in column 1, though we demonstrate in sensitivity analyses that the main results are robust to defining this variable in

other ways.[7] In terms of experience teaching Black students, each year, about 31 percent of white teachers are new to teaching a class that is at least 25 percent Black. Conditional on not being new to such a classroom, the average teacher has about 4 years of experience in classrooms that are at least 25 percent Black.

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' peers' characteristics and teaching experience. Specifically, we build on the models and identification strategies of Jackson and Bruegmann (2009) and Ost (2014) that we outline in sections 3.1 and 3.2, respectively.

3.1 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. We focus on same-grade peers because sociological research in education suggests that physical proximity is a key determinant of the likelihood that teachers have instructional advice-seeking ties (Spillane et al. 2017). Intuitively, the identification strategy pioneered by Jackson and Bruegmann (2009) 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 matters, as measured by experience and test score value-added, we similarly distinguish between having more and less effective Black peers using a set of mutually exclusive indicators. Importantly, we

[7] Specifically, we consider a definition based on years in classrooms that were at least 10% Black and a "years similar" measure motivated by Master et al. (2016) that measures the number of prior years in which the teacher taught a class with *at least* the same share of Black students (rounded to the nearest decile) as the current class.

parsimoniously control for the three types of teaching experience described in section 2.2: linearly and with an indicator for being "new" (Wiswall 2013).

Specifically, we estimate models of the form:

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

where i, j, g, s, and t index students, teachers, grades, schools, and years, respectively; 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; W includes classroom and time-varying teacher characteristics; and θ, ω, and φ represent grade-by-year, school-by-year, and teacher-by-school fixed effects (FE), respectively. f is a general function of the vector of the teacher's school-grade-year peers' average characteristics ($Peer$).[8] Otherwise, 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 fixed effects control for sorting into schools and differences across years in the EOG tests.

We estimate equation (1) using the full sample as well as the sample of white teachers' Black students using the FE estimator proposed by Correia (2016). We further stratify the data to estimate equation (1) separately by teachers' experience level, as relatively novice teachers may benefit more from the presence of a high-quality peer. The baseline estimates define novice as two or fewer years of experience, though we verify that results are robust to this arbitrary definition. 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, we verify 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).

The main innovation in Equation (1) 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

[8] When the outcome in a table is math scores, *Peer* includes the average of peers' math value added, and vice versa for ELA. For non-test score outcomes such as absences and suspensions we include both the math and ELA value added of peers. To be consistent within Tables that include a variety of outcomes, such as Tables 4 and 8, we include peer value added in both subjects. The math and reading results turn out to be robust to this decision.

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.

We report an exhaustive set of sensitivity analyses in Appendix A, which generally shows that estimates of equation (1) are robust to how the standard errors are clustered and to using Poisson regression for count outcomes. We also conduct two additional sensitivity analyses that are unique to the analysis of peer effects. First, we replace the Black peer indicator with a continuous measure of the percentage of peers who are Black. Second, we consider changing the cut point between novice and veteran teachers used in heterogeneity exercises.

3.2 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 (i.e., general human capital) and their experience teaching in the current grade (i.e., 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 FE and year-by-grade FE. We add a third type of experience to similar models that counts the years with a sizable share of Black students in the class 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. Our baseline models use experience in classrooms that were at least 25% Black, as per section 2.2.

Following Ost, we specify nonparametric specifications of all three experience types and estimate linear models that move *Peer* into *W* and are otherwise similar to equation (1):

$$y_{ijgst} = \beta X_{it} + \gamma W_{jt} + \theta_{gt} + \omega_{st} + \varphi_{js} + f(yrs) + g(same\ g) + h(25\%) + u_{ijgst}, \quad (2)$$

where the innovation is in *f*, *g*, and *h*, which represent general functions of each type of experience. Because we control for total years of teaching experience, our estimates of the

impact of years of teaching racially diverse classrooms will only be biased if students are sorted into classrooms based on teachers' diverse classroom experience conditional on a fixed level of overall teaching experience. Such sorting is unlikely, as Ost (2014) finds no evidence of sorting based on more easily observed and tracked same-grade experience.

Appendix B reports a series of sensitivity analyses that show that estimates of equation (2) are robust to three modeling and estimation choices in addition to the level at which standard errors are clustered. First, we consider alternative definitions of experience in racially diverse classrooms such as classrooms in which 10% of students are Black or classrooms that had at least the same decile share of Black students as the current classroom. Second, we change the point at which experience is top coded in the nonparametric specification of f, g, and h, as well as imposing linear functional forms. Finally, we estimate Poisson regression analogs to equation (2) for count outcomes such as absences and suspensions (Correia et al. 2020).

3.3 Sensitivity Analyses

Adjusting for fixed effects at the school-year level ensures 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 this 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. This also 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 (2009), we show that in equation (1) leads of the *Peer* variables (the characteristics of future peers) do not affect current performance.

4. Learning from Peers

4.1 Main Results

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 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 3 reports our baseline estimates of equation (1) using math EOG scores as the outcome.[9] 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 A1. In column 2 we augment this model to include an indicator for having at least one Black peer. The coefficient on this indicator is zero.

Following Jackson and Bruegmann (2009), in columns 3 and 4 of Table 3 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. Consistent with Jackson and Bruegmann (2009), the effect of peers' math VAM is 23% larger for novices than for veterans, though it is significant for both groups. By contrast, the effect of having at least one Black peer is an order of magnitude larger and is only statistically significant for novice teachers. Novice is defined as having 2 or fewer years of experience, as teachers improve markedly in their first couple of years (Bell et al. 2025); that said, Appendix Table A4 replicates column 3 of Table 3 using various definitions of novice and shows this result is robust to the choice of cutoff.

Panel B of Table 3 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 1.3% of a test-score SD, which amounts to a 3% 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 6% of a SD, which amounts to about a 12% increase. Interestingly, though, for this group the effect of peers' math value added shrinks

[9] Appendix Table A1 presents an analogous version for reading EOG scores. Appendix Table A2 shows that inference is robust to how standard errors are clustered. Appendix Table A3 uses the percentage of peers who are Black as the treatment of interest.

to zero. This suggests that for novice white teachers, effectively educating Black students requires a different set of specific skills that only a specific type of peer can provide. The opposite pattern appears in the sample of veteran white teachers studied in column 4: here the peer's race is irrelevant and what matters it their teaching productivity.[10] This suggests that there is a narrow window at the start of white teachers careers where whatever cultural competency, communication, and classroom management skills they learn from Black peers can be acquired.

Finally, Panels C and D of Table 3 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 indicators for having a veteran (as opposed to novice) Black Peer while novice is again defined as having two or fewer years of experience.[11] 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 comfortable or confident in giving 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 is highly effective, as measured by their math VAM, though both groups of Black peers significantly benefit their white colleagues. This is consistent with the results in Panel C, showing that some Black peers are more beneficial to their white colleagues' productivity educating Black students. Finally, the pattern observed in Panel B holds in Panels C and D regarding veteran white teachers: what matters is the teaching productivity, not race, of their peers.

[10] Importantly, these results are robust to how the Black peer treatment is defined. This is shown in Appendix Table A3, 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.

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

Table 4 re-estimates the peer-effect model 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, though throughout Table 4 the model controls for peers' math *and* reading value added. The effect of having at least one Black peer on novice white teachers' Black students' reading EOG scores is 4.3% of a SD, which resembles that on math EOG scores and represents a 10% increase. However, the estimate is only marginally significant.

Panels B and C of Table 4 investigate effects on student absences and suspensions, respectively.[12] Having a Black peer significantly reduces novice white teachers' Black students' absences. Column 2 of Panel B shows a reduction of about 0.6 absences, or a 12% reduction, while column 5 shows a 2 percentage point reduction in the chronic absence rate, which is a 42% 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%) reduction in the likelihood of ever being suspended from school. Together, Tables 3 and 4 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.

4.2 Falsification Exercises

Table 5 estimates the baseline peer-effects regressions from columns 2-4 of Table 3 for the white students of white teachers. If the results observed in Tables 3 and 4 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 this sample. It does not: the Black-peer indicator is not significant anywhere in Table 5 and in fact the estimates are fairly precisely estimated zeroes. This reinforces our interpretation of the peer effect finding and its validity: Black peers *specifically* 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

[12] Panel B of Appendix Table A5 replicates the absence and OSS-days results using a Poisson regression model.

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 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 6, 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, to provide a benchmark, Panel A reports the baseline model (with no leads or lags) estimated using the analogous leads / lags samples reported in Panel B. Columns 1 and 2 add one and two lags to the baseline model, respectively. Adding the lag does not change the estimated coefficient on the current Black-peer or peer math VAM inputs (Panel A), though the lagged peer math VAM measures are significant, suggesting a lasting impact of having an effective peer (again consistent with Jackson & Bruegmann 2009). The first lag of having a Black peer is statistically insignificant, but the second lag increases Black students' math EOG scores by about 2.3% of an SD, which closely resembles the contemporaneous effect of a Black peer for novice teachers seen in Table 3. This stickiness suggests that at least part of the Black peer effect 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 because teachers are necessarily younger when the twice-lagged peer was encountered, and we have seen that peer exposures are more impactful early in one's career.

Columns 3 and 4 of Table 6 add one and two leads to the baseline model, respectively. These leads are neither individually nor jointly 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 5, which showed no significant effects 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 6 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. And the lag results indicate that learning is a channel through which these peer effects operate.

5. Learning by Doing

In this section, we follow Ost (2014) in two notable respects. First, we restrict the analytic sample to new teacher cohorts, 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 7 reports our baseline estimates of equation (2) that document the impact of teaching experience on standardized end-of-grade (EOG) math scores. Appendix Table B1 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 7 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 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 (the 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 7 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, experience in classrooms that were at least 25% Black. 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.[13] 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 4% of a test-score SD when their teacher has had at least one year of experience in a diverse classroom.

Table 8 re-estimates the full model with all three types of experience shown in column 6 of Table 7 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 smaller than 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 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 B1). However, like in the case of EOG math scores, we see suggestive evidence that teachers' experience in racially diverse classrooms boosts Black students' EOG reading scores.

Columns 3 and 4 of Table 8 examine the effect of teacher experience on absences and chronic absence 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 and diverse-classroom experience attenuate these effects. It is puzzling that these skill-specific forms of experience reduce, sometimes significantly so, the total returns to experience. This could be the result of teachers who are new to a specific type of classroom environment focusing initially on socioemotional skills and then shifting more to academic skills as they become more comfortable

[13] Statistical inference in this and all tables presented in the main text is based on standard errors clustered at the teacher-year level. Appendix Tables B2, B3, and B4 show that the findings in Table 7 are robust to using a continuous measure of experience, changing the level of clustering of the standard errors, changing the non-parametric topcode of the experience variables, and changing the definition of experience in diverse classrooms.

in those classrooms. But this sort of puzzling result is exactly why our mixed-methods approach is valuable: qualitative data suggest another reason for this counterintuitive result. Regarding chronic absenteeism, it may be that white teachers in predominantly Black classrooms increasingly rely more on a specific flavor of discipline in ways that limit student-teacher relationships, at least for a subset of students, ultimately leading to more absenteeism (Lindsay et al. 2026). Indeed, student-teacher race match significantly reduces chronic absenteeism among Black students (Tran & Gershenson 2021).

Building on this discipline hypothesis, Columns 5 and 6 of Table 8 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, predicts exclusionary discipline (Hayes et al. 2023; Holt & Gershenson 2019; Lindsay & Hart 2017). Prior experience in racially diverse classrooms significantly reduces the number of OSS days though has no effect on the extensive margin.[14] This is consistent with teachers maintaining a better classroom environment but also forming better relationships with students that enable them to prevent disciplinary incidents from escalating to the point of a suspension. In sum, the results presented in Tables 7 and 8 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 boost the academic achievement of students from other backgrounds improves over time, and even more so with repeated exposure to diverse classrooms.

Finally, Table 9 estimates the baseline experience regression from column 6 of Table 7 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 9, we estimate the model for all teachers of all students in column 1 and for white students of white teachers in column 2; column 3 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,

[14] Appendix Table B5 reports Poisson regression analogs to columns 3 and 5 of Table 8. The Poisson estimates are qualitatively similar, suggesting that the results are robust to the use of a linear model.

communication, and classroom management skills that are orthogonal to racial competency.[15] 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.

The results in Table 9 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 50% larger than those accruing to their white students, as seen by comparing the estimates in column 2 to column 3. This suggests that white teachers learn a variety of broadly applicable skills from their experiences teaching in diverse classrooms, which are in addition to what they acquire via general teaching experience, some of which are unique to (or more impactful for) Black students.

6. Qualitative Analysis

6.1 Data and Method

The quantitative results presented in sections 4 and 5 provide evidence that teachers learn from past interactions with students and from their peers to more effectively teach Black students (as measured by end-of-grade math scores). Black peers also reduce absenteeism and suspensions among their white colleague's Black students. However, the experience results also present a puzzle: why does experience in diverse classrooms seemingly increase absenteeism among white teachers' Black students? While the quantitative estimates we present are well-aligned to the mechanisms we claim to isolate (skill-specific peer effects and learning-by-doing), we provide additional evidence on these mechanisms, and attempt to resolve the puzzling finding on absences, using qualitative data. We interviewed teachers to assess whether teachers' own words could help us to better understand our causal estimates.

While qualitative data rarely appears in economic research, it is not unprecedented (see Bewley (1995) for an early example). More recently, qualitative data have been combined with

[15] 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.

economic reasoning in several ways, including: to develop experimental interventions (Bergman et al. 2024), to generate hypotheses to be tested using larger-scale data sets (Bergman et al. 2024; DeLuca et al. 2023), to build theoretical models to rationalize puzzling behavior (DeLuca et al. 2024), and to inform the design of structural models that can be estimated and used in counterfactual policy analysis (DeLuca et al. 2026). In the present case, qualitative data have the specific and limited purpose of asking whether the group under study—teachers—directly and in their own words report information that aligns with the mechanisms that our econometric analyses suggest and, if so, can provide any additional suggestive information related to these mechanisms. Put simply, we ask if teachers say what we claim our estimates support. In general, our analysis of qualitative data corroborates our quantitative findings and provide novel information about the circumstances under which white teachers learn racial competency from their peers. Specifically, some white teachers emphasize that discussions about race are delicate and that they best learn from peers when the environment is non-judgmental. This finding has direct policy implications. Peer effects could presumably be stronger if colleagues were trained in how best to handle delicate subject matter.

We conducted interviews from a purposive sample of public-school teachers in three racially diverse counties in North Carolina between Spring 2022 and Spring 2023. Interviews were mainly focused on asking teachers to explain how they learn on the job, and specifically how they develop and improve their teaching practice both in general and specifically in racially diverse classrooms. We modeled our interview design on an approach known as "semi-structured narrative interviewing," a well-documented method in urban sociology that employs open-ended questions to elicit natural, long-form responses about complex behaviors and decision-making (e.g., Boyd & DeLuca 2017; DeLuca et al. 2016).[16] By avoiding overly specific probes, narrative interviewing promotes coverage of a broad set of topics in great detail, but does not push subjects towards specific sets or types of answers. 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 for failure to properly abide by a perceived

[16] Additional analyses using the qualitative data we describe here are provided in Lindsay et al. (2026), which also contains more detailed information on the survey instrument.

interviewing protocol. As a result, narrative interviewing yields rich data sets that can then be coded to test hypotheses.

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. Coding for the purpose of our 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 related 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 when teachers were novices, their school's racial and socioeconomic composition, experiences during the pandemic, and experiences with parents of students. 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.

6.2 Peer Results

Using the coded data, we first sought to determine whether white teachers report having learned from their Black peers, particularly regarding pedagogy and interaction with Black students. 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 their pedagogical development at least in part 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. Finally, 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 (within their first three years of service), which is consistent with our results on the importance of peer learning as teachers begin their careers. 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.

Peer learning occurred in a variety of settings. Interview data provide evidence of spillovers operating through more formal means, such as assigned mentorship programs, or through less formal interactions, such as casual conversations and friendship. One white female teacher explained how she developed (grew) as a teacher in diverse classrooms through her peer group that included Black teachers, saying:

> *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…*

Another teacher 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.*

In these two examples, it was a white teacher initiating the discussion and requesting feedback. But these interactions happen in the other direction too. For example, one white female teacher in a majority non-white school poignantly recounted being approached by some of her Black colleagues at the end of her second year:

> *I was pulled to the side by a couple Black teachers and they were like, you are being really mean to the Black kids and I was like, what? I broke down and … my initial reaction was, no I'm not! But then I'm like, … Am I mean to every Black kid? … I was trying to [communicate to students that] you're going to take me seriously- and I think it came across that I hated [the students], which is bad. I'm glad it happened because in my mind, I was just being a tough teacher and laying down the law and I don't think they would have pulled me aside and lied about it… and I think that's a big reason too of why I try so hard because I never want that to happen ever again.*

This quote is fascinating for a few reasons. First, it affirms the potentially long-lasting impacts of these peer effects seen in Table 6. Second, it also helps to explain our puzzling result regarding experience and absences seen in Table 8: in the absence of this intervention from her colleagues,

this teacher could have (unintentionally) ramped up her "laying down the law" approach in a way that caused students to disengage from her class and be absent more often, while simultaneously avoiding suspensions. Third, the Black teachers' intervention worked. They successfully communicated the disquieting information in a way that the teacher was able to digest and respond to in a productive way.

Another teacher discussed more formal settings for peer learning. One novice teacher described the following experience with her assigned mentor (of a different background):

> *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 [The 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.*

A noteworthy feature of all of these teachers' reports is not just that the presence of Black peers improved their practice. Rather, they expressed the value of non-judgmental peers to discuss difficult or sensitive issues. The latter teacher says she might "feel stupid for asking" and that "he didn't judge me." This point is of policy relevance as it suggests the importance of non-judgmental guidance if the objective is to improve white teachers' skill-specific human capital, as topics of race can be fraught. Moreover, this point would be difficult to observe in a large data and illustrates the value added by the qualitative interview data. Indeed, we find that of the 20 white teachers who discuss peer learning about Black students through Black colleagues, 57% recognize in some wat that the topic is delicate, conversations can be uncomfortable, or that an unjudgmental environment matters.

Taken together, the qualitative data affirm the quantitative results presented in section 4 and suggest the importance of peer learning through both formal and informal contact with Black peers. Being non-judgmental is particularly salient for informal peer learning in this domain. This suggests that if we could measure trustworthiness, coefficients capturing peer effects would be larger in environments with more trust and smaller in those without, consistent with the findings in Kraft and Papay (2014). As we cannot, we capture an average effect. Future work could explore this point further with additional data collection.

6.3 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:

> *"[In] 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 is important, especially when your teacher doesn't look like you."

The interviews also suggest an interplay between learning general and specific skills via trial and error over the course of one's career. For example, while addressing the issue of misbehavior in the classroom, one white female teacher in a majority-nonwhite classroom indicated that she developed an appreciation for a general teaching skill (being empathetic and developing trusting relationships with students) and a more specific skill of how develop and sustain those relationships with students who do not share the same background:

"It's all about the relationships because especially with the really difficult kids that have extreme behavioral problems, they'll work so much harder for you and behave so much better when they know that you care and have a strong relationship. So, that has been a really big thing for me that I've learned throughout the years is really… overpour with love for them…"

Another teacher explicitly addressed that the racial makeup of her school had been changing over time, which caused her to reassess not only her behaviors and approach to teaching but also empathizing with students to considering how they viewed her:

"I think in my years of teaching as the demographic has changed and evolved at my school and my awareness that students, even if I'm not being blatantly racist, students might already come in feeling a microaggression from me, for no fault of either of ours, but a child coming in and just looking at me and thinking, that teacher is different than me, not that it's already a shot against me, it's just I've got to work a little harder and I did not realize that that was already something that I needed to be working on. So, it's something that I have more awareness of now that I want to consciously work on."

These responses suggest a two-step mechanism through which teachers improve over time: they initially gain awareness of an issue or opportunity for improvement and then they actively work on getting better, either via trial and error or by seeking guidance from a colleague or mentor.

In sum, the qualitative data discussed in this section corroborate the findings from section 5 and provide insights into how exactly teachers benefit from experience accrued specific contexts, over and above the returns to general experience in the profession. Specific incidents or interactions cause teachers to stop and reflect on their methods and later to seek guidance on how to respond retroactively to the situation and/or be better prepared to handle similar incidents in the future. The resulting lessons stick with teachers for years.

7. Conclusion

This study provides evidence from North Carolina that white teachers learn how to more effectively teach Black 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), particularly early in their careers. These skill-specific peer effects are concentrated among newer teachers, though as veteran teachers generally productive peers become more important. The skill-specific experience provides a broader form of human capital that also benefits white teachers' white students, albeit to a lesser degree.

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 are 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 (Lindsay et al. 2026; Papay et al. 2020). In general, our findings provide practical guidance on how to use the existing teaching workforce 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. Earlier research has made the point that more productive peers tend to be more valuable than less productive peers and that experience drives learning and productivity gains. Our contribution is to consider both general and specific forms of both inputs in a setting where they can be measured simultaneously. We show that skill-specific peers and skill-specific experience matter over and above generally productive peers and general

experience. Moreover, these 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 more efficiently to maximize output both now and in the future.

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.

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

Bell, C., James, J., Taylor, E. S., & Wyckoff, J. (2025). Measuring returns to experience using supervisor ratings of observed performance: The case of classroom teachers. *Journal of Policy Analysis and Management*, *44*(1), 12-44.

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.

Bergman, P., Chetty, R., DeLuca, S., Hendren, N., Katz, L. F., & Palmer, C. (2024). Creating moves to opportunity: Experimental evidence on barriers to neighborhood choice. *American Economic Review*, *114*(5), 1281-1337.

Bewley, T. F. (1995). A depressed labor market as explained by participants. The American Economic Review, 85(2): 250-254.

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.

Callander, S. (2011). Searching and learning by trial and error. *American Economic Review*, *101*(6), 2277-2308.

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.

D'amico, D., Pawlewicz, R. J., Earley, P. M., & McGeehan, A. P. (2017). Where are all the Black teachers? Discrimination in the teacher labor market. *Harvard Educational Review*, *87*(1), 26-49.

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

Dee, T. S., & Penner, E. K. (2017). The causal effects of cultural relevance: Evidence from an ethnic studies curriculum. *American Educational Research Journal*, *54*(1), 127-166.

Delgado, W. (2025). Disparate teacher effects, comparative advantage, and match quality. *Economics of Education Review*, *106*, 102648.

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

DeLuca, S., Papageorge, N. W., Boselovic, J. L. (2024). Exploring the trade-off between surviving and thriving: Heterogeneous responses to adversity and disruptive events among disadvantaged black youth. *RSF: The Russell Sage Foundation Journal of the Social Sciences 10(1)*, 103-131

DeLuca, S., Papageorge, N. W., Gray, A., Gershenson, S., Boselovic, J. L., Sausedo, J. (2026). Using qualitative data to inform structural modeling: An application to post-secondary Education investments (No. w29472). National Bureau of Economic Research.

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. M., & Theobald, R. (2017). Does the match matter? Exploring whether student teaching experiences affect teacher effectiveness. *American Educational Research Journal*, *54*(2), 325-359.

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. (2022). *The Dreamkeepers: Successful teachers of African American children*. John Wiley & Sons.

Lazear, E. P. (2001). Educational production. *The Quarterly Journal of Economics*, 116(3), 777-803.

Levitt, S. D., List, J. A., & Syverson, C. (2013). Toward an understanding of learning by doing: Evidence from an automobile assembly plant. *Journal of Political Economy*, *121*(4), 643-681.

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. (2026). How Teachers Learn Racial Competency: The Role of Peers and Contexts. The Urban Review, 58(1), 10.

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.

Mas, A., & Moretti, E. (2009). Peers at work. *American Economic Review*, *99*(1), 112-145.

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.

Mincer, J. (1962). On-the-job training: Costs, returns, and some implications. *Journal of Political Economy*, *70*(5, Part 2), 50-79.

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), 359-388.

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.

Spillane, J. P., Shirrell, M., & Sweet, T. M. (2017). The Elephant in the Schoolhouse: The Role of Propinquity in School Staff Interactions about Teaching. *Sociology of Education*, 90(2), 149-171.

Tran, L., & Gershenson, S. (2021). Experimental estimates of the student attendance production function. *Educational Evaluation and Policy Analysis*, *43*(2), 183-199.

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	0.34	-0.44	-0.37
	(0.97)	(0.91)	(0.91)	(0.91)
Absences	5.22	5.49	4.73	4.67
	(6.83)	(6.97)	(6.96)	(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)	(0.70)	(1.66)	(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	1,731,363	889,770	107,892	314,657

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	0.41
Any Black or Hispanic Peer	0.32	0.19	0.66	0.42
Class Size	17.85	18.57	16.99	17.97
	(6.66)	(6.58)	(6.25)	(6.40)
Share Black Students	0.28	0.14	0.67	0.45
	(0.26)	(0.16)	(0.24)	(0.24)
Experience = 0 years	0.07	0.06	0.06	0.09
Experience | Experience > 0	12.13	12.41	13.95	11.45
	(9.01)	(8.81)	(9.89)	(9.08)
Same Grade Experience = 0 years	0.25	0.22	0.26	0.28
Same Grade Experience | Same Grade Experience > 0	4.31	4.46	4.29	4.11
	(3.38)	(3.40)	(3.24)	(3.29)
Yrs >25% Black = 0 years	0.46	0.59	0.18	0.31
Yrs >25% Black | Yrs >25% Black > 0	3.80	3.25	4.98	3.96
	(3.11)	(2.71)	(3.54)	(3.16)
Advanced Degree	0.32	0.32	0.35	0.30
Regular License	0.85	0.88	0.86	0.82
Certified	0.11	0.13	0.05	0.10
License Exam Score	0.06	0.17	-0.63	0.16
	(1.02)	(0.86)	(0.72)	(1.27)
Female	0.91	0.91	0.91	0.89
Math Value-added	0.03	0.05	-0.02	0.03
	(0.50)	(0.52)	(0.53)	(0.49)
ELA Value-added	0.02	0.03	0.02	0.02
	(0.45)	(0.47)	(0.49)	(0.44)
Peer Experience 0 years	0.07	0.06	0.09	0.08
Peer Experience | Peer Experience>0	11.24	11.69	11.08	11.00
	(6.29)	(6.20)	(7.01)	(6.38)
Peer Advanced Degree	0.32	0.32	0.31	0.31
Peer Regular License	0.84	0.87	0.79	0.82
Peer Certified	0.11	0.12	0.06	0.09
Peer License Exam Score	0.06	0.12	-0.19	0.03
	(0.65)	(0.61)	(0.57)	(0.84)
Peer Math Value-added	0.03	0.04	0.01	0.02
	(0.35)	(0.37)	(0.37)	(0.34)
Peer ELA Value-added	0.02	0.02	0.03	0.02
	(0.31)	(0.32)	(0.34)	(0.30)
N	110,761	50,246	8,112	24,074

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

Table 3. 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}		0.003	0.024**	0.002
		(0.003)	(0.011)	(0.003)
Peer math VAM	0.041***	0.041***	0.049**	0.040***
	(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}		0.013**	0.059***	0.004
		(0.006)	(0.022)	(0.006)
Peer math VAM	0.048***	0.049***	-0.007	0.048***
	(0.007)	(0.007)	(0.039)	(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	0.011*		0.080***	0.003
	(0.006)		(0.024)	(0.007)
Novice Black Peer	0.019**		0.001	0.011
	(0.009)		(0.037)	(0.011)
Peer math VAM	0.049***		-0.009	0.047***
	(0.007)		(0.039)	(0.008)
D. Peer Effectiveness, for White Teachers' Black Students				
More Effective Black Peer	0.038**		0.146**	0.027
	(0.016)		(0.068)	(0.018)
Less Effective Black Peer	0.012**		0.054**	0.003
	(0.006)		(0.023)	(0.006)
Peer math VAM	0.047***		-0.026	0.045***
	(0.007)		(0.039)	(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 4. 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}	0.013**	0.059***	0.004	0.011*	0.043*	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	-0.49	-0.36	-0.35	-0.43	-0.32
		Annual Absences			1{Chronically Absent}	
1{Black peer}	-0.104*	-0.596***	-0.076	-0.005**	-0.021***	-0.004
	(0.057)	(0.210)	(0.067)	(0.002)	(0.008)	(0.002)
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}	0.008	-0.113**	0.002	-0.005**	-0.020*	-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 5. Teacher Peer Effects on EOG Math Scores for White students of White teachers

	White students of White teachers		
	All	Novice	Veteran
	(1)	(2)	(3)
1{Black peer}	-0.004	-0.001	-0.004
	(0.004)	(0.020)	(0.005)
Peer math VAM	0.042***	0.054	0.040***
	(0.004)	(0.035)	(0.004)
N	889,771	142,395	746,812
E(y|no Black peer)	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 6. 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}	0.016**	0.003	0.015*	0.011	0.009
	(0.008)	(0.010)	(0.008)	(0.012)	(0.011)
Peer math VAM	0.046***	0.039***	0.043***	0.034***	0.038***
	(0.009)	(0.011)	(0.010)	(0.013)	(0.012)
B. Leads and Lags					
1{Black peer}	0.016**	0.004	0.016*	0.010	0.008
	(0.008)	(0.010)	(0.008)	(0.012)	(0.011)
Lag 1{Black peer}	-0.003	0.006			-0.001
	(0.007)	(0.010)			(0.010)
Second Lag 1{Black peer}		0.023**			
		(0.009)			
Lead 1{Black peer}			-0.005	0.006	0.001
			(0.008)	(0.012)	(0.011)
Second Lead 1{Black peer}				-0.009	
				(0.011)	
Peer math VAM	0.040***	0.034***	0.041***	0.033**	0.031***
	(0.009)	(0.012)	(0.010)	(0.013)	(0.012)
Lag Peer math VAM	0.026***	0.015			0.025**
	(0.008)	(0.011)			(0.011)
Second Lag Peer math VAM		0.026***			
		(0.010)			
Lead Peer math VAM			0.009	-0.001	0.008
			(0.010)	(0.014)	(0.012)
Second Lead Peer math VAM				-0.004	
				(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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 7. Effects of Teaching Experience on EOG Math Scores

Sample:	All Student & Teachers		Black Students of White Teachers			
	(1)	(2)	(3)	(4)	(5)	(6)
1 year	0.038***	0.030***	0.037***	0.027***	0.026***	0.023***
	(0.004)	(0.004)	(0.008)	(0.008)	(0.008)	(0.009)
2 years	0.056***	0.045***	0.065***	0.053***	0.050***	0.047***
	(0.005)	(0.005)	(0.010)	(0.011)	(0.011)	(0.011)
3+ years	0.070***	0.055***	0.069***	0.050***	0.048***	0.041***
	(0.005)	(0.006)	(0.012)	(0.012)	(0.012)	(0.012)
1 yr same grade		0.015***		0.018***		0.013***
		(0.002)		(0.004)		(0.005)
2 yrs same grade		0.022***		0.022***		0.015**
		(0.002)		(0.005)		(0.006)
3+ yrs same grade		0.027***		0.035***		0.028***
		(0.003)		(0.006)		(0.007)
1 yr >25% Black					0.022***	0.014**
					(0.005)	(0.006)
2 yrs >25% Black					0.033***	0.023***
					(0.007)	(0.008)
3+ yrs >25% Black					0.041***	0.026***
					(0.009)	(0.009)
N	1,606,055		292,107			
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 (to category) teacher). >25% Black refers to the number of years the teacher has previously taught in a classroom in which at least 25% of students were Black. 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. In each regression, each set of three experience-type indicators are jointly statistically significant at the 99% confidence level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 8. 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	0.023***	0.009	-0.039	-0.006*	-0.059**	-0.003
	(0.009)	(0.009)	(0.093)	(0.003)	(0.023)	(0.004)
2 years	0.047***	0.012	-0.213*	-0.013***	-0.045	-0.006
	(0.011)	(0.012)	(0.123)	(0.005)	(0.031)	(0.006)
3+ years	0.041***	0.038***	-0.322**	-0.017***	-0.061*	-0.010
	(0.012)	(0.013)	(0.137)	(0.005)	(0.032)	(0.006)
1 yr same grade	0.013***	0.004	0.061	0.000	-0.002	-0.004*
	(0.005)	(0.005)	(0.052)	(0.002)	(0.011)	(0.002)
2 yrs same grade	0.015**	-0.009	0.040	0.001	0.022	0.002
	(0.006)	(0.006)	(0.065)	(0.002)	(0.015)	(0.003)
3+ yrs same grade	0.028***	-0.003	0.139*	0.005*	0.011	0.002
	(0.007)	(0.007)	(0.073)	(0.003)	(0.016)	(0.003)
1 yr >25% Black	0.014**	0.004	0.026	0.005**	-0.022	0.002
	(0.006)	(0.006)	(0.061)	(0.002)	(0.015)	(0.003)
2 yrs >25% Black	0.023***	0.017**	0.049	0.005*	-0.059***	-0.004
	(0.008)	(0.008)	(0.083)	(0.003)	(0.021)	(0.004)
3+ yrs >25% Black	0.026***	0.011	0.187*	0.010***	-0.032	0.002
	(0.009)	(0.010)	(0.101)	(0.004)	(0.025)	(0.005)
Joint sig. (p value)	0.005	0.090	0.229	0.028	0.013	0.054
N	292,107	291,592	292,365	292,365	279,240	279,240
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). >25% Black refers to the number of years the teacher has previously taught in a classroom in which at least 25% of students were Black. 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. The joint significant test p values are for an F test of the joint significance of the three >25% indictors. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 9. Effects of Teaching Experience on EOG Math Scores for Different Subgroups

Sample:	All	White S White T	Black S White T
	(1)	(2)	(3)
1 year	0.027***	0.028***	0.023***
	(0.004)	(0.006)	(0.009)
2 years	0.042***	0.042***	0.047***
	(0.005)	(0.007)	(0.011)
3+ years	0.051***	0.056***	0.041***
	(0.006)	(0.008)	(0.012)
1 yr same grade	0.013***	0.014***	0.013***
	(0.002)	(0.003)	(0.005)
2 yrs same grade	0.019***	0.020***	0.015**
	(0.003)	(0.003)	(0.006)
3+ yrs same grade	0.025***	0.026***	0.028***
	(0.003)	(0.004)	(0.007)
1 yr >25% Black	0.010***	0.014***	0.014**
	(0.003)	(0.004)	(0.006)
2 yrs >25% Black	0.016***	0.017***	0.023***
	(0.004)	(0.006)	(0.008)
3+ yrs >25% Black	0.014***	0.015**	0.026***
	(0.005)	(0.007)	(0.009)
Joint sig. (p value)	0.0001	0.002	0.005
N	1,606,055	819,157	292,107
E(y|exp = 0)	-0.15	0.16	-0.57

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). >25% Black refers to the number of years the teacher has previously taught in a classroom in which at least 25% of students were Black. 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 A: Sensitivity Analyses & Additional Results for Peer Effects

Table A1. 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	0.013	0.002
		(0.002)	(0.010)	(0.003)
Peer read VAM	0.023***	0.023***	0.016	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}		0.010*	0.043*	0.004
		(0.006)	(0.024)	(0.007)
Peer read VAM	0.030***	0.030***	0.042	0.025***
	(0.007)	(0.007)	(0.046)	(0.008)
N	313,983		70,592	242,211
E(y|no Black peer)	-0.35		-0.43	-0.32

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.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A2. 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}		0.003	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	0.041	0.041	0.049	0.040
Teacher-year	(0.003)	(0.003)	(0.024)	(0.004)
Teacher	(0.004)	(0.004)	(0.028)	(0.004)
School-year	(0.004)	(0.004)	(0.027)	(0.004)
School	(0.006)	(0.006)	(0.030)	(0.006)
T-year & Student	(0.003)	(0.003)	(0.024)	(0.004)
T & S	(0.004)	(0.004)	(0.028)	(0.004)
Sch-yr & Student	(0.004)	(0.004)	(0.027)	(0.004)
School & Student	(0.006)	(0.006)	(0.030)	(0.006)
B. White Teachers' Black Students				
1{Black peer}		0.013**	0.059	0.004
Teacher-year		(0.006)	(0.022)	(0.006)
Teacher		(0.006)	(0.028)	(0.007)
School-year		(0.006)	(0.026)	(0.007)
School		(0.007)	(0.034)	(0.009)
T-year & Student		(0.006)	(0.022)	(0.006)
Teacher & Student		(0.006)	(0.028)	(0.007)
Sch-yr & Student		(0.006)	(0.026)	(0.007)
School & Student		(0.007)	(0.034)	(0.009)
Peer math VAM	0.048***	0.049***	-0.007	0.048***
Teacher-year	(0.007)	(0.007)	(0.041)	(0.008)
Teacher	(0.008)	(0.008)	(0.047)	(0.009)
School-year	(0.008)	(0.008)	(0.045)	(0.009)
School	(0.010)	(0.010)	(0.053)	(0.011)
T-year & Student	(0.007)	(0.007)	(0.040)	(0.008)
T & S	(0.008)	(0.008)	(0.047)	(0.009)
Sch-yr & Student	(0.008)	(0.008)	(0.046)	(0.009)
School & Student	(0.010)	(0.010)	(0.053)	(0.011)

Note: This table is identical to Panels A and B of Table 3, with seven additional standard errors clustered at successively higher levels (or two-way clustered by student as well).

Table A3. Teacher Peer Effects on EOG Math Scores

Teacher Experience	All	All	Novice	Veteran
	(1)	(2)	(3)	(4)
A. All Students				
% Black Peers		0.002	0.034	0.001
		(0.005)	(0.026)	(0.006)
Peer math VAM	0.041***	0.041***	0.049**	0.040***
	(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		0.009	0.085	-0.005
		(0.011)	(0.055)	(0.013)
Peer math VAM	0.048***	0.048***	-0.007	0.047***
	(0.007)	(0.007)	(0.039)	(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 3 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A4. 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}	0.031	0.024	0.024**	0.015*	0.015**	0.010
	(0.052)	(0.017)	(0.011)	(0.008)	(0.007)	(0.006)
Peer math VAM	0.081	0.032	0.049**	0.057***	0.058***	0.053***
	(0.092)	(0.038)	(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}	0.004	0.037	0.059***	0.038**	0.049***	0.047***
	(0.116)	(0.032)	(0.022)	(0.018)	(0.015)	(0.012)
Peer math VAM	-0.207	0.014	-0.007	-0.004	0.024	0.040**
	(0.202)	(0.105)	(0.039)	(0.029)	(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 3 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A5. Poisson Regressions for Count Outcomes

Teachers	All	Novice	Veteran	All	Novice	Veteran
Outcome		Annual Absences			OSS Days	
	(1)	(2)	(3)	(4)	(5)	(6)
1{Black peer}	-0.024**	-0.160***	-0.021	-0.033	-0.789***	-0.052
	(0.012)	(0.044)	(0.013)	(0.072)	(0.289)	(0.091)
Peer Math VAM	-0.011	-0.233**	-0.002	-0.159	-0.916	-0.144
	(0.017)	(0.109)	(0.018)	(0.108)	(0.730)	(0.129)
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: The regression models in this table are the same as those in columns 1-3 of Table 4 for the count outcomes annual absences and out of school suspensions (OSS), but specified as Poisson rather than linear regressions. 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Appendix B: Sensitivity Analyses & Additional Results for Returns to Experience

Table B1. Effects of Teaching Experience on EOG Reading Scores

Sample:	All		Black Students of White Teachers			
	(1)	(2)	(3)	(4)	(5)	(6)
1 year	0.011***	0.007**	0.013*	0.010	0.010	0.009
	(0.003)	(0.004)	(0.008)	(0.009)	(0.009)	(0.009)
2 years	0.021***	0.017***	0.016	0.016	0.011	0.012
	(0.005)	(0.005)	(0.011)	(0.011)	(0.012)	(0.012)
3+ years	0.028***	0.023***	0.043***	0.042***	0.037***	0.037***
	(0.005)	(0.005)	(0.012)	(0.013)	(0.013)	(0.013)
1 yr same grade		0.007***		0.005		0.004
		(0.002)		(0.004)		(0.005)
2 yrs same grade		0.008***		-0.003		-0.008
		(0.002)		(0.006)		(0.006)
3+ yrs same grade		0.009***		0.001		-0.003
		(0.002)		(0.006)		(0.007)
1 yr >25% Black					0.006	0.004
					(0.005)	(0.006)
2 yrs >25% Black					0.013*	0.017**
					(0.007)	(0.008)
3+ yrs >25% Black					0.009	0.011
					(0.009)	(0.010)
N	1,603,619		291,592			
E(y|exp = 0)	-0.146		-0.520			

Notes: The outcome is standardized End-of-Grade (EOG) reading 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). >25% Black refers to the number of years the teacher has previously taught in a classroom in which at least 25% of students were Black. 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 B2. Effects of Experience on EOG Math Scores: Sensitivity to how Standard Errors are Clustered

Sample:	All			Black Students of White Teachers		
	(1)	(2)	(3)	(4)	(5)	(6)
Experience	-0.003	-0.003	-0.002	-0.002	-0.002	-0.003
	(0.002)	(0.002)	(0.004)	(0.004)	(0.004)	(0.004)
1{New to NC}	-0.041	-0.030	-0.040	-0.030	-0.029	-0.025
Teacher-year	(0.004)	(0.004)	(0.008)	(0.008)	(0.008)	(0.008)
Teacher	(0.004)	(0.005)	(0.009)	(0.010)	(0.010)	(0.010)
School-year	(0.004)	(0.004)	(0.009)	(0.009)	(0.009)	(0.009)
School	(0.005)	(0.006)	(0.010)	(0.011)	(0.011)	(0.011)
T-year & Student	(0.004)	(0.004)	(0.008)	(0.008)	(0.008)	(0.008)
T & S	(0.004)	(0.005)	(0.009)	(0.010)	(0.010)	(0.010)
Sch-yr & Student	(0.004)	(0.004)	(0.009)	(0.009)	(0.009)	(0.009)
School & Student	(0.005)	(0.006)	(0.010)	(0.011)	(0.011)	(0.011)
Same grade Exp.		0.001		0.004		0.004
		(0.001)		(0.001)		(0.001)
1{New to grade}		-0.018		-0.016		-0.010
Teacher-year		(0.002)		(0.004)		(0.005)
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)		(0.005)
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)
Yrs >25% Black					0.009	0.007
					(0.003)	(0.003)
1{New to >25%B}					-0.021	-0.015
Teacher-year					(0.005)	(0.006)
Teacher					(0.006)	(0.007)
School-year					(0.006)	(0.006)
School					(0.007)	(0.008)
T-year & Student					(0.005)	(0.006)
T & S					(0.006)	(0.006)
Sch-yr & Student					(0.005)	(0.006)
School & Student					(0.007)	(0.007)
N	1,606,055			292,107		
E(y|exp = 0)	-0.154			-0.572		

Notes: This table is analogous to Table 7 in the main text with one exception: the nonparametric experience indicators are replaced by continuous measures of each type of experience 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 >25% Black refers to the number of years the teacher has previously taught in a classroom in which at least 25% of students were Black. 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.

Table B3. Effects of Teaching Experience on EOG Math Scores

"Diverse" Coding:	>10% Black		Years Similar	
	(1)	(2)	(3)	(4)
1 year	0.020**	0.020**	0.033***	0.025***
	(0.009)	(0.009)	(0.008)	(0.008)
2 years	0.041***	0.041***	0.058***	0.050***
	(0.011)	(0.011)	(0.010)	(0.011)
3+ years	0.037***	0.035***	0.059***	0.046***
	(0.013)	(0.013)	(0.012)	(0.012)
1 yr same grade		0.010*		0.016***
		(0.005)		(0.004)
2 yrs same grade		0.008		0.018***
		(0.006)		(0.006)
3+ yrs same grade		0.023***		0.031***
		(0.007)		(0.006)
1 yr diverse	0.026***	0.019***	0.011***	0.007*
	(0.005)	(0.006)	(0.004)	(0.004)
2 yrs diverse	0.042***	0.035***	0.022***	0.018***
	(0.007)	(0.008)	(0.005)	(0.005)
3+ yrs diverse	0.049***	0.033***	0.019***	0.015***
	(0.009)	(0.010)	(0.005)	(0.005)

Notes: The outcome is standardized End-of-Grade (EOG) math scores. Samples contain self-contained 4th and 5th grade classrooms and are restricted to the Black students of white teachers who entered North Carolina Public Schools in 2001 or later. N = 302,474 and the control group mean = -0.572. The omitted reference group for all experience types is 0 years (i.e., a new (to category) teacher). Diverse coding: >10% Black refers to the number of years the teacher has previously taught in a classroom in which at least 10% of students were Black. Years 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. In each regression, each set of three experience-type indicators are jointly statistically significant at the 99% confidence level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B4. Effects of Teaching Experience on EOG Math Scores: Sensitivity to Nonparametric Specification

Max Bin	2	3	4	5	6	7	8	9	10
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Total 1	0.022***	0.023***	0.024***	0.024***	0.025***	0.025***	0.025***	0.025***	0.025***
	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Total 2	0.044***	0.047***	0.050***	0.051***	0.051***	0.052***	0.052***	0.052***	0.051***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Total 3		0.041***	0.041***	0.043***	0.044***	0.044***	0.045***	0.045***	0.044***
		(0.012)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Total 4			0.053***	0.049***	0.051***	0.051***	0.052***	0.052***	0.051***
			(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Total 5				0.063***	0.062***	0.063***	0.064***	0.064***	0.062***
				(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Total 6					0.069***	0.071***	0.073***	0.073***	0.071***
					(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Total 7						0.069***	0.074***	0.075***	0.072***
						(0.017)	(0.017)	(0.017)	(0.017)
Total 8							0.069***	0.076***	0.073***
							(0.018)	(0.018)	(0.018)
Total 9								0.063***	0.066***
								(0.019)	(0.019)
Total 10									0.054***
									(0.020)
Same 1	0.013***	0.013***	0.013***	0.013***	0.013***	0.013***	0.013***	0.014***	0.014***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Same 2	0.020***	0.015**	0.015**	0.015**	0.015**	0.015**	0.015***	0.016***	0.016***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Same 3		0.028***	0.020***	0.019***	0.019***	0.019***	0.020***	0.021***	0.021***
		(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Same 4			0.037***	0.039***	0.039***	0.039***	0.041***	0.041***	0.041***
			(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Same 5				0.032***	0.028***	0.028***	0.030***	0.030***	0.030***
				(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Same 6					0.035***	0.034***	0.036***	0.037***	0.037***
					(0.009)	(0.009)	(0.009)	(0.009)	(0.010)
Same 7						0.037***	0.031***	0.032***	0.032***
						(0.010)	(0.010)	(0.010)	(0.010)
Same 8							0.049***	0.045***	0.045***
							(0.011)	(0.012)	(0.012)
Same 9								0.051***	0.053***
								(0.013)	(0.013)
Same 10									0.048***
									(0.014)
>25% 1	0.013***	0.013***	0.013***	0.013***	0.013***	0.013***	0.013***	0.014***	0.014***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
>25% 2	0.020***	0.015**	0.015**	0.015**	0.015**	0.015**	0.015***	0.016***	0.016***

(Table B4, continued from previous page)

	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
>25% 3		0.028***	0.020***	0.019***	0.019***	0.019***	0.020***	0.021***	0.021***
		(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
>25% 4			0.037***	0.039***	0.039***	0.039***	0.041***	0.041***	0.041***
			(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
>25% 5				0.032***	0.028***	0.028***	0.030***	0.030***	0.030***
				(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
>25% 6					0.035***	0.034***	0.036***	0.037***	0.037***
					(0.009)	(0.009)	(0.009)	(0.009)	(0.010)
>25% 7						0.037***	0.031***	0.032***	0.032***
						(0.010)	(0.010)	(0.010)	(0.010)
>25% 8							0.049***	0.045***	0.045***
							(0.011)	(0.012)	(0.012)
>25% 9								0.051***	0.053***
								(0.013)	(0.013)
>25% 10									0.048***
									(0.014)

Notes: The outcome is standardized End-of-Grade (EOG) math scores. Samples contain self-contained 4th and 5th grade classrooms and are restricted to the Black students of white teachers who entered North Carolina Public Schools in 2001 or later. N = 302,474 and the control group mean = -0.572. The omitted reference group for all experience types is 0 years (i.e., a new (to category) teacher). The categories are total years of teaching experience (Total), years of experience teaching in the current grade (Same), and years the teacher has previously taught in a classroom in which at least 25% of students were Black (>25%). In each column, the "max bin" includes teachers with that level of experience, *or more* (i.e., column 2 replicates the main specification from Table 7 of the main text, in which Total 3 includes teachers with 3 or more years of total teaching experience). 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 B5. Poisson Regressions for Count Outcomes

Teachers	Absences	OSS Days
	(1)	(2)
1 year	-0.023	-0.199**
	(0.017)	(0.092)
2 years	-0.058***	-0.141
	(0.022)	(0.122)
3+ years	-0.077***	-0.170
	(0.025)	(0.135)
1 yr same grade	0.019**	-0.054
	(0.009)	(0.059)
2 yrs same grade	0.012	0.091
	(0.012)	(0.069)
3+ yrs same grade	0.038***	0.001
	(0.013)	(0.080)
1 yr >25% Black	0.004	-0.073
	(0.011)	(0.069)
2 yrs >25% Black	0.005	-0.323***
	(0.015)	(0.091)
3+ yrs >25% Black	0.023	-0.215**
	(0.019)	(0.107)
N	225,649	123,486
E(y|exp = 0)	6.9	0.80

Notes: Samples are restricted to the Black students of white teachers. The omitted reference group for all experience types is 0 years (i.e., a new teacher). Yrs >25% Black refers to the number of years the teacher has previously taught in a classroom in which at least 25% of students were Black. 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.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.