ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 23-831

OK Boomer: Generational Differences in Teacher Quality

Nhu Nguyen
University of Chicago
Inclusive Economy Lab

Ben Ost
University of Illinois at
Chicago

Javaeria A. Qureshi
University of Illinois at
Chicago

We document that recent generations of elementary school teachers are significantly more effective in raising student test scores than those from earlier generations. Measuring teachers' value-added for Black and white students separately, the improvements in teaching for Black students are significantly larger than those seen for white students. The race-specific improvements in teacher quality are driven by white teachers. Analyses of mechanisms suggest that changing teachers' biases may be one potential channel. Our results suggest reason for optimism since these teacher quality differences should lead to improved student learning and a narrowing of the Black-white test score gap over time.

VERSION: August 2023

Suggested citation: Nguyen, Nhu, Ben Ost, and Javaeria Qureshi. (2023). OK Boomer: Generational Differences in Teacher Quality. (EdWorkingPaper: 23-831). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/xfeb-ca25

OK Boomer: Generational Differences in Teacher Quality

Nhu Nguyen, Ben Ost, and Javaeria A. Qureshi*

June 2023

We document that recent generations of elementary school teachers are significantly more effective in raising student test scores than those from earlier generations. Measuring teachers' value-added for Black and white students separately, the improvements in teaching for Black students are significantly larger than those seen for white students. The race-specific improvements in teacher quality are driven by white teachers. Analyses of mechanisms suggest that changing teachers' biases may be one potential channel. Our results suggest reason for optimism since these teacher quality differences should lead to improved student learning and a narrowing of the Black-white test score gap over time.

JEL codes: I24, H75, J24, J15, J16

*Nguyen: University of Chicago Inclusive Economy Lab; Ost: Department of Economics, University of Illinois at Chicago; Qureshi: Department of Economics, University of Illinois at Chicago, javaeria@uic.edu
We thank Elizabeth Cascio, Joshua Goodman, Caroline Hoxby, Dan Kreisman, Steve Rivkin, Jonah Rockoff, Doug Staiger, Jake Vigdor, seminar participants at Boston University Wheelock College, Columbia University Teacher's College, Dartmouth College, the Federal Reserve Bank of Chicago, Illinois State, Lahore University of Management Sciences, the NBER Education meeting, the University of Illinois at Chicago, Wellesley College, Williams College, the Institute for Research and Poverty Summer Research Conference, the Association for Education Finance and Policy (AEFP) Conference, the Association for Public Policy and Management (APPAM) Conference, the Southern Economic Association (SEA) Conference, and the E-Con of Education seminar for helpful comments and suggestions. We are grateful to the North Carolina Education Research Data Center for providing us access to the data for this project. This data can be obtained by submitting a data application directly with NCERDC (https://childandfamilypolicy.duke.edu/research/nc-education-data-center/application-forms/). The authors declare that they have no relevant or material financial interests that relate to the research described in this paper. IRB approval (Protocol #2017-1134) was obtained for this project.

1. Introduction

Changes in the role of women and in race relations are two of the most fundamental transformations of American society over the past 70 years. Researchers are increasingly interested in the implications of this reduced discrimination for economic productivity. It is against this backdrop that we explore cohort differences in worker productivity. This study focuses on teachers, a class of workers for whom there is a direct measure of productivity available, so we do not have to infer productivity from wages (e.g., Hsieh, Hurst, Jones, and Klenow 2019). Though most occupations were affected by the social changes related to gender and race, the implications for teaching are potentially far-reaching because it has a predominantly female and white workforce teaching an increasingly diverse student body.[1] This study provides the first evidence on birth cohort differences in overall and race-specific teacher productivity using value-added measures that assess overall effectiveness at raising student test scores as well as effectiveness for raising test scores for Black and white students separately.

Traditionally, educated women were limited primarily to teaching and health-related professions, but their options have widened considerably with the gender desegregation of the labor market since 1960 (Goldin, 1990; Black and Juhn 2000). It is often thought that earlier cohorts of teachers were particularly good teachers because those women entered the profession when there were not many outside labor market opportunities for women (e.g. Temin 2002). A New York Times article describes Baby Boomer teachers as "enduring teachers from an older

[1] According to the National Center for Education Statistics, seventy-six percent of US public school teachers were women, and about eight in ten teachers identified as non-Hispanic white compared to 47% of all US public school students in the 2017-2018 school year.

era of higher quality."[2] Yet, there is no evidence on how teacher quality varies across birth cohorts.

The latter half of the twentieth century was also one of great social change with respect to race relations. The civil rights era affected people of all age groups, but the changing social norms were acutely important for the generations that came of age after these norms had changed. Thus, even today, there are significant differences in racial attitudes across generations with more recent generations self-reporting less racist views (Bobo et al. 2012). These changes have the potential to affect teacher productivity since effective teaching relies heavily on interpersonal connections and past work shows that teacher expectations directly affect student achievement (Rosenthal and Jacobson 1968).

Though these broad social forces motivate our investigation of generational differences in teacher productivity, there have also been other changes that could potentially affect teachers' overall and race-specific productivity such as changes in teachers' own educational experiences, teacher training, and certification requirements. For example, culturally relevant pedagogy became popular in the early 1990s so recent generations of teachers are more likely to have been trained on the importance of inclusive instruction compared to earlier generations.

This paper uses administrative records from North Carolina public schools for 1997-2016 to estimate birth cohort differences in overall and race-specific teacher quality or value-added.[3] Our study is related to Nagler, Piopiunik and West (2020) who assess how value-added varies with business cycle conditions at career start and we largely follow their two-step empirical approach. First, we estimate each teacher's permanent value-added following the extensive

---

[2] Levy, Harold O. "Why the Best Don't Teach" New York Times, September 9, 2000.
[3] Teachers are valued for much more than their effectiveness in raising student test scores. For example, Jackson (2018) shows that teacher effects on non-test score outcomes also have effects on long run outcomes. We follow the literature in referring to test score value-added as "teacher quality".

literature on teacher quality.[4] The second step examines how these value-added measures evolve across the 1946-1992 birth cohorts. We also use generation indicators (Baby Boomers, Generation X, Millennials) as a convenient way to aggregate changes across birth cohorts.[5]

In some regards, our analysis is a straightforward, descriptive exercise. However, teacher value-added must be estimated and here we face a challenge related to the classic age-cohort-year problem. Teachers from different birth cohorts are observed with different levels of experience in our data. Since student achievement improves with teacher experience, our value-added estimation must control for these differences in experience to identify the permanent component of teacher quality. Accounting for experience is not straightforward because, as noted by Rockoff (2004), experience and year are typically perfectly colinear within a teacher. We address this by imposing a cap to the returns to experience, following an extensive literature that shows that though teachers improve considerably in the early years, this improvement does not continue indefinitely. If teachers continue to improve after the point at which we impose a flat experience profile, we will inadvertently attribute this unexplained improvement to their value-added estimate and teachers from earlier cohorts who have higher experience will appear to have higher permanent value-added. The data supports the cap we impose, and we also estimate value-added measures with higher caps. Importantly, the results are similar if we restrict the analysis to highly experienced teachers where the influence of how we model the effect of experience is more limited.

[4] Rothstein (2010) raises concerns about the validity of value-added estimates. However, multiple studies have found that value-added measures are unbiased predictors of teachers' impacts on student achievement following random assignment or quasi-experimental changes in teacher assignment (e.g., Kane and Staiger 2008; Kane et al. 2013; Chetty, Friedman and Rockoff 2014a; Bacher-Hicks, Kane and Staiger, 2014). Teacher value-added has also been shown to be predictive of long term outcomes such as earnings (Chetty, Friedman and Rockoff 2014b).

[5] According to the Pew Research Center, Baby Boomers were born between 1946-1964, Gen X were born between 1965-1980, and Millennials were born after 1980.

We document two main results. First, teacher quality has improved significantly across birth cohorts. Compared to a Baby Boomer teacher, assignment to a Gen X teacher raises a student's math score by 0.027 standard deviations (SD) and assignment to a Millennial teacher raises it by 0.049 SD. Second, recent cohorts of teachers are significantly better at teaching both Black and white students, but the improvements for Black students are significantly larger than those for white students. A Black student assigned to a Millennial teacher has a math score 0.09 SD higher than when assigned to a Baby Boomer teacher. In contrast to these significant birth cohort differences in teacher effectiveness for math, we find little evidence of differences in either overall or race-specific teacher effectiveness at raising reading scores across cohorts.[6]

The generational improvements in teaching Black students are sizable and a similar order of magnitude to estimates of same-race teacher matching effects (e.g., Goldhaber and Hansen 2010). Assessing generational differences in quality for white and Black teachers separately, we find that the differential improvement in teaching Black versus white students is driven by white teachers. The race-specific improvement for Black teachers is more muted but recent generations of Black teachers appear to be substantially more effective than Baby Boomer Black teachers.

We investigate the role of several mechanisms that might contribute to the cohort differences in teacher quality. There is little evidence that the cohort differences are driven by differential sample selection across cohorts.[7] We also show that the relationship between teacher quality and the likelihood of persisting in teaching is unchanged across cohorts. Similar to past

[6] This pattern of results echoes that in many past studies which find smaller or no effects of school inputs on reading scores compared to math scores. Many researchers suspect this reflects that math is almost exclusively learned at school, but reading is influenced by the family and outside factors (Jackson, Rockoff, Staiger 2014).

[7] Differences between the earliest and more recent cohorts could reflect sample selection since we only observe early cohorts if they persist to the years covered by our data. We find suggestive evidence against this explanation as the results are similar if we focus on recent cohort teachers who are more comparable to the early cohorts in their eventual persistence in the profession. Importantly, sample selection is only a potential mechanism for trends for the earliest cohorts because starting with the mid-1970s birth cohorts, we observe teachers from career start.

studies which find that observed characteristics do not predict differences in teacher effectiveness (e.g. Jackson, Rockoff, Staiger), we find that teacher characteristics are unable to explain the generational differences in overall quality.

We find some evidence that recent generations of teachers exhibit lower racial evaluation bias, i.e., a smaller racial gap in their *expectations* for the performance of their Black and white students with the same blindly scored performance. This difference in teachers' racial evaluation bias is unlikely to fully explain the cohort differences in teacher quality, but it is suggestive of tangible generational differences across teachers in how they view their students. Using data from the Project Implicit Database, we also show that recent cohorts of teachers have significantly lower levels of implicit racial bias. The generational differences in bias, combined with the fact that the race-specific improvements in quality are driven by white teachers, suggests that changing teachers' biases are one plausible mechanism for the race-specific improvements in teacher quality we find.

This study contributes to three literatures. First, it relates to a nascent literature highlighting the significant costs of discrimination for economic productivity (e.g., Cook 2014, Hsieh et al. 2019, Aizer et al. 2020, Aneja and Xu, 2022). Our focus on teachers allows us to have direct measures of productivity and our ability to study a race-specific measure of productivity is quite novel in this space. Second, this study adds to the literature on the effects of teacher demographics on student achievement which has studied the role of teacher race and gender extensively but ignored birth cohort (see Redding (2019) for a review). To the best of our knowledge, this study is the first to analyze the effect of teacher birth cohort on student outcomes. Lastly, studies have examined how selection into teaching has changed based on teachers' academic aptitude (Corcoran et al., 2004; Goldhaber and Walch, 2013; Gitomer 2007;

Bacolod 2007).[8] Test scores of entering teachers declined from the 1940s to the 1970s birth cohorts (Corcoran et al. 2004), then rose from the 1970s to the late 1980s and 1990s birth cohorts (Goldhaber and Walch 2013; Lankford et al. 2014). Yet, measures of academic aptitude have been found to be poor predictors of teacher quality (Clotfelter et al. 2007; Harris and Sass 2011; Jackson, Rockoff and Staiger 2014). We add to this work by documenting the evolution of teacher quality across birth cohorts directly instead of using teacher attributes that are at best weakly correlated with teacher quality. Our focus on value-added also allows us to assess changes in quality for Black and white students separately, which is infeasible when inferring teacher quality from factors like their own test scores.

In addition to providing useful description, understanding cohort differences in teacher quality is important for forming expectations of how teacher quality is evolving. Our results suggest reason for optimism regarding the impending retirements of earlier cohorts of teachers since this should naturally lead to an improvement in student learning and a narrowing of the Black-white test score gap over time. Since teacher quality is the most important school input, aggregate changes in teacher quality have significant implications for students' education and long-term outcomes (Chetty, Friedman and Rockoff 2014b; Rivkin, Hanushek and Kain 2005).

2. Conceptual framework

This section provides a conceptual discussion of the factors that could contribute to differences in teacher quality across cohorts, some of which are expected to lead to different patterns for Black and white students. We describe changes in the teachers' own formative and educational

[8] This work has mostly focused on how credentials of prospective teachers have changed over time. Since we study teachers who persist to our data's time period (1997-2016), our analysis captures a different snapshot in the teacher pipeline. This distinction may be important since global forces (e.g., increased labor market opportunities for women) that shape the composition of entering teachers may simultaneously affect the attrition of earlier cohorts.

experiences, changes in teacher training programs and certification requirements, followed by a discussion of how teachers' occupational choices have evolved in recent decades.

2.1. Changes in Racial Attitudes and Educational Experiences

Social changes tied to the civil rights movement have led to significant changes in racial attitudes, with recent generations having more progressive attitudes with regards to race. For example, in a 2002-2003 Pew survey, 21% of Baby Boomer individuals disagreed with the statement "it is alright for Blacks and whites to date each other", whereas 12% of Gen X and 8% of Millennials disagreed with it (Taylor, Funk and Craighill 2006). Attitudes of teachers follow a similar trend: in Section 6.3, we show that, much like non-teachers, recent cohorts of teachers have significantly lower levels of implicit racial bias than earlier cohorts of teachers.

Many Baby Boomer teachers in the South would have attended K-12 schools that were fully segregated whereas almost all Gen X individuals would start kindergarten in integrated schools (Cascio et al. 2008, 2010).[9] Johnson (2011) shows that school desegregation led to significant improvements in per-pupil spending for Blacks, so one might expect generational differences across Black teachers due to these changes in K-12 spending.[10] Although desegregation did not affect school inputs or educational outcomes for whites (Johnson 2011), they gained Black peers in their schools. White teachers attending school after desegregation might have better attitudes and communication, given that increased exposure to minority groups lowers prejudice, and leads to more comfortable interaction with the minority groups.[11]

2.2. Changes in Teacher Licensure and Teacher Training

[9] Racial integration in K-12 schools rose steadily until 1980 and has declined modestly since then (Clotfelter, Ladd, Vigdor 2006). Hinrichs 2016 shows that the average white person in the South attended a college with only 3% of Blacks in 1968 compared to 15% in 2011.
[10] School desegregation improved Black students' educational attainment (Guryan, 2004; Reber, 2010; Hanushek et al., 2009), adult incomes (Ashenfelter et al., 2006), and decreased criminal offending (Weiner, Lutz, Ludwig, 2009).
[11] This has been documented by Van Laar et al. 2005, Boisjoly et al. 2006, Sidanius et al. 2008, and Burns, Corno and La Ferrara, 2015 using randomized exposure to minorities in college.

There were several changes in teacher licensure standards over this time. From 1926 to 1964, North Carolina certification was solely based on college course work, but a licensure exam requirement was added in 1964 (Thompson 1979).[12] In 1967, this exam was expanded to include pedagogy exams, and additional requirements for certification were adopted along with increased salaries under the Elementary and Secondary School Reform Act of 1983 (Darling-Hammond et al. 2022). Though licensure requirements can alter the composition of who enters teaching, the requirements are generally based on factors with little correlation with value-added.

Teachers from different cohorts are trained at different times and the topics emphasized in teacher training programs have changed over time.[13] These shifts in curriculum have followed from developments in pedagogical theory with a general movement from more positivist to more constructionist pedagogy. For example, math teacher training in the 1950s-1970s emphasized procedures to arrive at the correct answer whereas in the 1980s it emphasized the importance of conceptual understanding and process (Kilpatrick 1992). Along with a greater emphasis on student-centered learning, the late 1980s saw a structural shift where ideas like culturally relevant pedagogy (CRP), Gardner's multiple intelligences, and cooperative learning became popular. CRP, in particular, could lead to differential effects by student race and ethnicity as it emphasizes matching content and instructional practice to each student's cultural context (Milner

[12] When first established, the standardized exam performance standards were binding for many teachers with a starkly different pattern for Black and White teachers. Over half of Black prospective teachers scored below the certification threshold whereas fewer than 5% of White teachers did so.

[13] Though the content of teacher training programs varies by institution and instructor, there are many common elements across programs and the modal curriculum has shifted over time as beliefs regarding pedagogical best practices evolved. These broader changes may affect teachers from all cohorts if experienced teachers engage in professional development or learn from their newly minted peers. The cohort component exists to the extent that there is a persistent effect to the intensive initial training period.

2011).[14] It also emphasizes the importance of having and communicating high expectations of all students.

It is also possible that recent cohorts of teachers are more adept at teaching the current curriculum, and more familiar with technology and standardized testing. These factors could contribute to differences in teacher effectiveness for students overall, but it is not clear that they would lead to differences in race-specific teacher effectiveness.

2.3 The Role of Occupational Choices and Hiring Decisions

The labor market changes over this period could have changed selection into teaching. The occupational choices of prospective teachers can be understood in the context of a Roy model where individuals choose their occupation based on the return to skills in each sector along with the correlation between skill for teaching and skill outside of teaching. Bacolod (2007) uses this model to analyze the theoretical effects of lower barriers to entry outside of teaching and increased pay compression within teaching, both of which are relevant for the time period we study (Hoxby and Leigh 2004). Bacolod shows that the overall effect of these changes on teacher quality is theoretically ambiguous. Individuals who remain in teaching following the expansion in non-teaching opportunities could be negatively selected if the correlation between teaching and non-teaching skill is high. Alternatively, those who remain in teaching despite the growth in outside opportunities might be particularly well-suited to teaching if skills are more multidimensional.

Regardless of the form of self-selection into teaching, the implication of this self-selection for who actually becomes a teacher depends on which teachers schools select. If

[14] In a review of textbooks from the 1960s, Allen (1971) concludes "In too many instances, the Black child doesn't see anyone of his own color in positions other than the unskilled". Relatedly, Edmonds (2022) finds that black students perform better with HBCU-trained teachers who are trained in culturally-fluent pedagogical practices that are conducive to Black student academic success.

schools always hire the highest value-added teachers in a given pool, then any reduction in the size of the pool should mechanically at least weakly reduce teacher quality. Alternatively, if schools hire teachers based on credentials that are not strongly predictive of value-added, changes in self-selection may have a limited effect on the value-added of hired teachers. Jacob et al. (2018) and Hinrichs (2021) suggest that schools do not necessarily hire the applicants with the best credentials and given that value-added is difficult to measure ex-ante, it is likely that schools do not consistently select applicants with the highest value-added. Of course, not every teacher who applies for and is hired by a school will remain in teaching for the rest of her career. The same forces that affect the entry of teachers can impact their exit decisions, and, therefore, attrition from teaching may also be differential across birth cohorts.

3. Empirical Strategy

Our empirical strategy to assess cohort differences in teacher quality follows the Nagler et al. (2020) two-step approach. The first step estimates the permanent component of value-added for each teacher and the second step characterizes how this permanent value-added differs across teachers from different birth cohorts.

To estimate each teacher's value-added, we follow the extensive literature that estimates value-added by predicting test scores in year *t* as a function of test scores in year *t-1* along with other covariates.[15]

$$A_{ijgt} = \gamma A_{i(t-1)} + X_{it}\beta + f(\mathrm{Exp_{jt}}) + \theta_j + \lambda_g + \delta_t + \varepsilon_{ijgt} \quad (1)$$

[15] Kane and Staiger (2008) show that using the lagged model, they are able to replicate experimental estimates and Chetty, Friedman and Rockoff (2014a) argues that controlling for lagged scores is sufficient to account for sorting of students to teachers. Rothstein (2017) argues against this conclusion. Most recently, Bacher-Hicks et al. (2019) presents new experimental validation and summarizes the literature evaluating bias in the value-added models: a value-added model that controls for lagged scores has been consistently validated across seven studies, three quasi-experimental and four experimental ones.

A_{ijgt} is the test score for student *i* in teacher *j*'s class, in grade g, and year *t*. $A_{i(t-1)}$ is student *i*'s lagged test score, X_{it} are a set of student covariates including race, ethnicity, gender, as well as indicators for special education and limited English proficiency status. The vector $f(\mathrm{Exp_{jt}})$ controls for teacher *j*'s experience in year *t* in a flexible manner discussed below, as well as an indicator for whether the teacher is eligible for retirement. The vector θ_j are the teacher fixed effects or value-added measures which capture each teacher's permanent, time-invariant teaching quality, i.e., her effectiveness at raising student test scores. The teacher quality measure θ_j is net of the effect of the other controls including experience.[16] Since teacher's birth year is a fixed characteristic for each teacher, the teacher quality measures θ_j will include the effect of birth cohort and we will analyze how the teacher quality measures vary with birth cohort in step two.

Since a key question of interest is how teacher quality varies by student race, in addition to estimating Equation (1) across all students, we also estimate it separately for a teacher's Black students and her white students. These three regressions yield three estimates of teacher quality for each teacher: θ_j (defined above), θ_j^B(teacher *j*'s quality for teaching Black students) and θ_j^W (teacher *j*'s quality for teaching white students). As suggested in McCaffrey et al. (2012), we impose a sum to zero condition on these value-added estimates so that each estimate is relative to the grand mean rather than to an arbitrary hold-out teacher. The sum to zero condition means that θ_j , $\theta_j^B$and θ_j^W all have zero mean by construction.

In the second step, we relate these teacher quality measures to the teachers' birth cohort indicators. In addition to showing how these measures vary across each of the 46 birth cohorts from 1946-1992, we use generation categories to conveniently summarize differences across

[16] In certain contexts, researchers use Empirical Bayes shrinkage to improve the precision of value-added estimates at the expense of introducing bias. We average value-added across many teachers to characterize cohort change, so the improvement in precision is of secondary importance.

birth cohorts. For the overall teacher quality measure, θ_j, we simply estimate the following teacher-level regression where Baby Boomers are the omitted group.

$$\theta_j = \beta_0 + \beta_1 Gen\,X_j + \beta_2 Millennial_j + \epsilon_j \quad (2)$$

The coefficient β_1 captures the average difference in teacher quality between Gen X and Baby Boomer teachers whereas β_2 captures the average difference between Millennial and Baby Boomer teachers.[17] To assess how race-specific teacher quality varies across generations, we stack the θ_j^B and θ_j^W estimates so that each teacher appears in the data twice. Defining the stacked vector as θ_{jb}, and defining R_b as an indicator for whether the teacher quality observation corresponds to value-added for teaching Black students or white students, we estimate

$$\theta_{jb} = \delta_0 + \delta_1 R_b + \delta_2 Gen\,X_j + \delta_3 Mill_j + \delta_4 (Gen\,X_j * R_b) + \delta_5 (Mill_j * R_b) + \epsilon_j \quad (3)$$

The δ_2 and δ_3 coefficients capture the average difference in teacher quality for teaching white students (the omitted group) across generations. The coefficients δ_4 and δ_5 capture the difference in the generational quality gradient for teaching Black students relative to the generational quality gradient for teaching white students. We cluster standard errors at the teacher level.

Equation (1) attributes any school-level effects to the teachers at that school. For example, if well-resourced schools or effective principals raise test scores directly, this will be reflected as higher value-added for each teacher in that school. This is a concern for our estimation if teachers from different birth cohorts sort towards different schools. To assess

[17] Since teacher quality is estimated, there is measurement error in the dependent variable, which leads to larger standard errors compared to if we observed true teacher value-added.

whether cross-school differences drive our results, we modify Equation (1) to estimate teacher-by-school fixed effects θ_{jsb}, and use these as the dependent variable in the following equation: [18]

$$\theta_{jsb} = \gamma_0 + \gamma_1 R_b + \gamma_2 Gen\,X_j + \gamma_3 Mill_j + \gamma_4 (Gen\,X_j * R_b) + \gamma_5 (Mill_j * R_b) + R_b * \delta_s + \epsilon_j \quad (4)$$

Equation (4) includes $R_b * \delta_s$, i.e., race-by-school fixed effects, to ensure that the generational differences in race-specific teacher quality are identified within the same schools. It is important to note that the within-school comparison of teacher quality also alters the interpretation of the parameters so that these estimates could differ from our preferred specification, even if all estimates are unbiased. For example, if Baby Boomers tend to teach at schools with higher quality peers, their relative position within their school may be lower than Millennials, even if they have the same absolute quality.

Since teachers from different birth cohorts are observed with different experience in our data, it is critical that our value-added estimation correctly account for the effect of experience. A key challenge is that in models that control for teacher fixed effects, experience and year are perfectly colinear for teachers with continuous careers (Rockoff 2004). This is closely related to the classic age-cohort-year problem and there are three approaches used in the literature to address it. Most commonly, researchers follow Rockoff (2004) and impose a "flat part" of the experience profile, where teachers are assumed to not improve after a certain experience threshold. Papay and Kraft (2015) suggest an alternative two-step approach that allows for estimating year and experience effects flexibly but requires the assumption that cohort effects are

[18] We replace the teacher fixed effects from Equation (1) with teacher-by-school fixed effects because teachers may teach in more than one school. Estimating a separate teacher-by-school fixed effect for teaching spells at different schools in step 1 allows us to assign each teacher quality estimate to a unique school, *s*, and include a race-by-school fixed effect in Equation (4).

zero.[19] Finally, Wiswall (2013) notes that it is possible to flexibly account for year and experience non-parametrically (or linearly) by utilizing gaps in teaching to help break the collinearity between experience and year. This approach relies on the assumption that leaves are exogenous and that skills do not depreciate during the leaves from teaching, which we are reluctant to do given recent work showing that teacher quality depreciates with time away from teaching (Dinerstein, Megalokonomou, and Yannelis, 2022).

Our preferred strategy follows the first approach by imposing the assumption that teachers stop improving after a certain experience level. This is a functional form assumption, motivated by the well-documented finding that teachers improve considerably in their early years on the job but these improvements do not continue indefinitely. Early studies typically set this threshold around 5 years, but Papay and Kraft argue that there are returns to teaching past 10 years and note that models that impose a "flat portion" have the potential to understate the returns to experience throughout. Estimating value-added while imposing a cap to experience returns attributes any difference between the true experience returns and the estimated returns to the estimate of permanent value-added. Because earlier cohorts of teachers generally have more experience than recent ones, imposing an experience cap in a range where teachers are still improving might bias us towards finding that earlier cohorts of teachers have higher permanent quality than recent cohorts. Our preferred approach assumes that there are no further returns to experience after 15 years of experience, i.e., $f(\mathrm{Exp}_{\mathrm{jt}})$ includes indicators for the teacher's experience equal to 1, 2, …, 14, and ≥ 15 years. The experience profile in our context supports

[19] More precisely, they assume that the stock of teacher quality conditional on experience is constant over the time period they study (2002-2009). Papay and Kraft provide evidence in support of this assumption for the anonymous district they study and emphasize that the assumption becomes harder to satisfy with a longer panel. Our panel is nearly 20 years long and imposing a constant stock of quality would be at odds with our goal of estimating cohort change.

imposing this cap on the returns to experience, and we show that our conclusions are robust to using higher caps as well.

Teachers from different cohorts may also have differences in non-teaching experience that we do not observe. For example, if a lawyer decides to become a teacher at age 50, she will have zero years of teaching experience, but her experience as a lawyer may make her more effective than the typical novice teacher. Because these "second career" teachers are more likely to come from earlier cohorts, including these teachers has the potential to overestimate the experience-adjusted quality of early cohorts. As such, we drop teachers where the teacher's experience level, year of BA and calendar year suggest that the teacher did not begin teaching until later in life. Specifically, we drop any teacher whose age minus experience is not within 5 years of when the teacher obtained a BA.[20] We have also verified robustness to specifications that control for age minus teaching experience minus year of BA, which effectively shuts down identifying variation from the timing of entry into teaching.

4. Data and Analysis Sample

This study uses administrative data from North Carolina public schools for the years 1997-2016 that were obtained from the North Carolina Education Research Data Center (NCERDC). The data has information on students, classrooms, teachers, and schools, and includes course membership information required to match students to teachers from 2007 onwards. Before 2007, the data includes the grade taught by each teacher and the exams she proctored, but there is no direct measure of which teacher taught which students. For elementary classrooms, the proctor is almost always the classroom teacher, and institutional factors suggest that it is very

[20] Some "second career" teachers might obtain a BA for the first time at an older age, so we also require that age at BA is 25 or less.

unlikely that a teacher proctoring students in her assigned grade is not their classroom teacher.[21] We follow Clotfelter, Ladd and Vigdor (2007), Rothstein (2010), Ost (2014), and Jackson and Bruegmann (2009) in imposing some sample restrictions, which further increase our confidence that the students are matched to their classroom teachers. Specifically, we consider a proctor to be the classroom teacher as long as the teacher's grade assignment matches the grade of the proctored exam, the proctor primarily administers tests in that grade, and the classroom has more than 10 and fewer than 40 students. On average, our teacher value-added estimates are based on approximately 74 students and race-specific value-added estimates are based on approximately 46 white and 21 black students, respectively.

Students in North Carolina are tested annually in reading and math in grades 3-8. We study elementary school teachers because in middle school, students are often tracked according to ability, which makes identifying teacher quality more challenging. At the elementary school level, students are typically in self-contained classrooms where the same teacher teaches reading and math. Since our analysis requires a valid test score and lagged test score, we focus on students enrolled in grades four and five. We standardize test scores by grade and year prior to making sample restrictions so that there is a mean of zero and a standard deviation of one in the population. We drop any children who are repeating a grade because grade repeaters are likely to have atypical achievement gains. Appendix Table A1 shows descriptive statistics for our estimation sample.[22] Just over half the students are white, 26% are Black, and 8% are Hispanic.

---

[21] Jackson and Bruegmann (2009) note that "tests are always administered by the students' own teachers when these teachers are present. Also, all students in the same grade take the exam at the same time; thus, any teacher teaching a given subject in a given grade will almost certainly be administering the exam only to her own students".

[22] We have slightly different samples when analyzing math versus reading achievement. Appendix Table A1 shows summary statistics for the math sample and the results are very similar for the reading sample. Our analysis sample is slightly advantaged compared to the general population since requiring non-missing lagged scores leads to selecting a more stable group of students.

The data has longitudinal information on teachers including demographic characteristics, baccalaureate institutions, degrees earned, and completed years of experience. Table 1 shows summary statistics for teachers separately by generation. About 90% of teachers are female and this has not changed much across the generations. In contrast, the fraction of teachers who are Black fell substantially from 18 percent among the Baby Boomers to 8 percent among Millennials, mirroring a trend that is seen nationwide. There are very few Hispanic teachers and other race teachers in North Carolina. As expected, recent generations are observed with fewer years of experience and are less likely to be observed with an advanced degree.

Gen X and Millennial teachers are less likely to have received their BA from a southern institution compared to Baby Boomer teachers, perhaps suggesting that fewer of the recent generation teachers grew up in the South (which would be in line with recent in-migration trends for North Carolina). We assess changes in the selectivity of college attended by merging with IPEDS data on undergraduate institutions' 2010 Carnegie classification and find that recent generations of teachers attended more selective colleges.[23] The fraction of teachers who attended low selectivity colleges falls from 28% among Baby Boomers to 12% among Millennials. Appendix Figure A1 includes trends in the college selectivity measures separately by teacher race, showing that the improvements in selectivity of college attended are driven almost exclusively by Black teachers and that these changes for Black teachers are large. Almost 83% of Black Baby Boomer teachers attended a low selectivity college compared to 47% of Millennial Black teachers, whereas the fraction of Black teachers attending medium selectivity colleges increased from 12% to 37% across these cohorts.

5. **Results**

[23] The measures of college selectivity are described in the notes for Table 1.

Figure 1 plots the average teacher quality, θ_j, in math and reading against teachers' year of birth. The birth year means of teacher quality in math are generally flat for the Boomer cohorts, then they rise appreciably among the Gen X cohorts before levelling off at a higher point for the Millennials. In contrast to math, the mean teacher quality estimates for reading show little evidence of systematic variation across birth year, except for perhaps a small decline among the most recent cohorts.

To summarize these differences across birth cohorts and to assess their statistical significance, we present estimates of Equation (2) regressing teacher quality against generation indicators. Column 1 of Table 2 shows that assignment to a Gen X teacher raises students' math scores by 0.027 SD and assignment to a Millennial teacher raises math scores by 0.049 SD versus being taught by a Baby Boomer, respectively.[24] Mirroring the results in Figure 1b, Column 4 shows that assignment to a Gen X teacher instead of a Baby Boomer teacher has no effect on reading scores, whereas assignment to a Millennial teacher instead of a Baby Boomer teacher lowers reading scores by 0.009 SD.

Figures 2a and 2b plot the birth year means of the estimates of each teacher's race-specific teaching quality obtained by estimating Equation (1) separately for Black students and white students. The black dots indicate birth year averages of teacher quality for teaching Black students and the grey dots indicate birth year averages of teacher quality for teaching white students. Figure 2a shows a pattern of increasing teacher quality across birth cohorts for teaching both Black students and white students. Strikingly, the improvement in teacher quality we see for teaching Black students is much stronger than that seen for teaching white students. Since the value-added estimates for Black students and those for white students are separately normalized

[24] Value-added is measured in student standard deviations – the same units as the standardized exam.

to have mean zero, we can only infer how average teacher quality trended for Black students versus for white students from this figure: in other words, this exercise is not informative about level differences in teacher quality. Interestingly, much of the improvement in teacher quality starts with the early Gen X cohorts for teaching both white students and Black students which then continues through the Millennial cohorts for teaching Black students but largely stops for white students with the end of the Gen X cohort. In contrast to these results for math, Figure 2b shows that there is no evidence of changing teacher quality in reading across teacher birth cohort for either Black or white students.

Columns 3 and 6 of Table 2 present results of Equation (3) estimating generational differences in race-specific value-added focusing on black and white students. The coefficients on the main effects of the generation indicators quantify the generational differences in teacher quality for white students. Assignment of a white student to a Gen X or Millennial teacher versus a Baby Boomer teacher raises math scores significantly by 0.027 SD and 0.046 SD, respectively. The coefficients on the interaction terms of teacher generation with the indicator for whether the value-added is for Black students show that the generational improvements in teacher quality for teaching Black students are significantly larger than those for teaching white students. Summing up the main effects and interaction coefficients shows that a Black student assigned to a Gen X teacher has a math score 0.057 SD higher than when assigned to a Baby Boomer teacher and a Black student assigned to a Millennial teacher has a math score 0.09 SD higher than when assigned to a Baby Boomer teacher. Though these results are presented relative to Baby Boomer teachers, the implied improvements in teacher quality between Gen X and Millennial teachers for white students (0.019 SD) and Black students (0.033 SD) in Column 3 are also statistically significant. For reference, Columns (2) and (5) show generational differences in overall quality from Equation (2) for the sample in Columns (3) and (6), i.e. excluding students who are non-black and non-white, and restricting to teachers with value-added estimates for both white students and black students.[25]

To put these generational differences in perspective, we compare them to the effects of teacher experience and the same-race teacher matching effect. Papay and Kraft (2015) find that the total returns to teacher experience for math range from roughly 0.1 to 0.15 SD. Using North Carolina data, Goldhaber and Hansen (2010) estimates that black students perform 0.04 SD better when assigned to a Black rather than white teacher. The Gen X – Baby Boomer gap in teacher quality for Black students we find is similar to the same-race teacher effect, whereas the Millennial – Baby Boomer gap in teacher quality is larger than the same-race teacher effect and almost as large as the total returns to teaching experience. Interestingly, the Millennial – Baby Boomer gap in relative effectiveness for teaching Black versus white students (0.044) is similar to the race-matching estimates (0.04) from Goldhaber and Hansen (2010).

Studies in education commonly find that school inputs or interventions have no or smaller effects on reading compared to math. Studies have also typically found larger variance in teacher effects on math than in reading. This is thought to reflect the fact that schools account for the vast majority of learning in math, but parents and other outside factors play a larger role in reading and language skills (Jackson, Rockoff, Staiger 2014). Since there is no evidence of generational differences in teacher quality when it comes to raising reading scores for either Black or white students, we focus on teacher quality in math going forward.

[25] In results not shown, we estimate kernel density plots to assess generational differences across the distribution of teacher value-added. We find that the generational improvements in value-added are broad based and occur throughout the distribution of effectiveness.

Table 3 shows results from Equation (4) that estimate within-school differences in teacher quality. As described in Section 3, to compare teachers within the same schools only, we alter the first stage to estimate teacher-by-school effects instead of teacher effects so that teachers who teach in more than one school appear multiple times in our second stage, leading to a larger sample than in our preferred specification. Column 1 replicates our preferred specification for convenience. Column 2 shows an intermediate specification that estimates teacher-by-school fixed effects in the first step but does not impose a within-school comparison in the second step, yielding qualitatively similar results to our preferred specification. In Column 3, we add race-by-school fixed effects to the regression, which summarizes the generational differences in teacher quality that exist within schools. In Column 3, the estimated main effects are smaller, the interaction terms are slightly larger, but the overall takeaway is largely unchanged.

5.1. Accounting for teacher experience

Our preferred approach overcomes the collinearity between experience and year by imposing a cap to the returns to experience at 15 years of experience. Here we provide empirical evidence in support for this assumption and show that our findings are robust to relaxing this assumption. First, in Figure 3 we plot the estimated experience profile under various assumptions regarding the flat portion. Specifically, we estimate specifications that include indicators out to N years of experience and then include an indicator for N or more years of experience. We vary the cap, starting with 10 and going up to 30 in increments of 5.

As in past studies, we find the biggest returns to teaching experience in the first few years of teaching. Consistent with Papay and Kraft (2015), the returns to experience continue beyond these first few years and we observe positive returns even after 10 years of teaching.[26] That said, there appear to be negligible returns to experience past 15 years of teaching and the exact placement of the cap (15, 20, 25 or 30) does not substantially alter the experience profile, indicating that a cap at 15 years captures the functional form well.[27] Relatedly, if we begin with our preferred model and add in indicators for having experience over 20 years and over 25 years, the coefficients on these two additional indicators are approximately zero and the p-value testing their joint significance is 0.99. Additionally, we examine how our estimates of the generational differences in teacher quality change if we estimate value-added estimates from alternative models where we modify the experience cap to 20, 25 or 30 years instead. Appendix Table A2 shows that the pattern of generational effects is qualitatively similar regardless of the experience cap. A remaining potential concern is that teachers' effectiveness might start to decline later in their career and we would disproportionately capture earlier cohort teachers at this stage. To assess this possibility, the last column of Table A2 estimates generational differences in teacher quality using our preferred specification but restricts the sample to observations where the teacher has no more than 30 years of teaching experience. The estimated generational differences are similar overall, suggesting that potential late career declines in teaching effectiveness are not driving our findings.

Finally, we focus exclusively on observations where the teachers have already accumulated substantial experience to assess whether our parametrization of the effect of experience drives the generational pattern we find. Given the absence of returns to experience

[26] When estimating the same set of models, Papay and Kraft (2015) find similar total returns to experience (ranging from 0.1 to 0.15 SD), though they find that capping experience at 15 instead of 20 yields different returns estimates. In our sample, the estimates of the total returns are very similar when we impose the cap at 15 versus 20.

[27] In theory, these results could also be consistent with perfectly linear growth above 15 years of experience, but this scenario is unlikely since diminishing returns to experience are expected at some point.

past 15 years, generational differences in quality among teachers with more than 15 years of experience are unlikely to be driven by experience differences. In Column 5 of Appendix Table A3, we show the difference between Gen X and Baby Boomers, focusing only on observations where the teacher has more than 15 years of experience. Though we lose the ability to study Millennial teachers who have not yet accumulated 15 years of experience, we find it reassuring that the Baby Boomer-Gen X gap in teacher quality in Column 5 is very similar to that in our preferred specification.[28]

Restricting the sample to teachers with more than 15 years of experience avoids the need to control for experience, but it limits the cohorts we can study so we also assess how estimates differ if we focus on teacher-year observations that have at least 10, 8 or 5 years of experience. In these specifications, we still control for experience (capping at 15) since there are returns expected in these ranges. The idea is that since the majority of the returns to experience accrue in the first few years, focusing on more experienced observations will reduce the influence of the way in which we model experience. Columns 2-4 of Appendix Table A3 show that the estimated generational gaps are quite similar as we vary the experience requirement.

6. Exploration of Mechanisms

There are a variety of potential mechanisms that could drive the generational differences in teacher quality that we find. This section empirically investigates the role of survival, teacher characteristics, and teacher biases in the generational patterns we find.

6.1 The role of survival

[28] Column 5 covers a slightly different sample of Gen X teachers compared to Column 1 because the last two Gen X cohorts (1979-1980) have not accumulated 15 years of experience in our data. We have verified that the results are similar if we instead require 13+ years of experience so that all Gen X cohorts can be included.

The reduced form generational differences in teacher quality we document reflect differences across teachers who persist to our data's time period (1997-2016). As such, they reflect permanent differences in teacher quality across generations as well as the effects of differential selection and retention in teaching across the generations. For the Baby Boomers, we only observe experienced teachers so there is no way to separately identify initial selection from subsequent retention. All Millennial teachers and most Gen X teachers start teaching after the start of our data so we can assess whether there is differential retention of later versus earlier birth cohorts. Understanding how selection out of teaching is changing across cohorts is of direct interest and our study provides the first evidence on this question.

Specifically, we study the likelihood of a teacher leaving without teaching for five years in North Carolina public schools. Because there is no direct measure of teacher exit in the data, it has to be inferred from no longer seeing a teacher in the database.[29] To avoid censoring issues and to allow for the possibility of leaves of absence, we restrict attention to teachers who start as novices between 1997 and 2010 and characterize a teacher as leaving without teaching for five years if a teacher fails to accumulate 5 years of experience within 7 years of starting.[30]

Figure 4 shows the likelihood of teachers leaving without teaching for five years separately for teachers with high VA (math value-added higher than the median) and low VA (math value-added lower than the median) for those born in the years 1970-1987. We see that more recent cohorts have lower attrition rates than earlier ones, but this pattern is similar for high

[29] The NC public school database includes teachers of all grades and subjects. Teachers who are absent from the database may have left teaching for another occupation, for non-market work, or they may have started teaching in a private school or at a public school outside the state of North Carolina. Though some of the teachers who leave NC public schools may still be teaching, assessing exit from NC public schools is the relevant margin for understanding the evolution of teacher quality in our data.

[30] Since not all teachers who fail to accumulate 5 years of experience within 7 years may have necessarily left teaching, we have verified that results are similar if we allow 10 years for the teacher to accumulate 5 years allowing for more leaves of absence. Results are also similar if we allow for no leaves of absence.

and low value-added teachers.[31] There appears to be slightly improved selective attrition of teachers among the later cohorts, though the magnitude of the change is small and not statistically significant.[32] Furthermore, we see similar cohort patterns for value-added for white students and value-added for Black students in Appendix Figure A2. Changing selective attrition of teachers across cohorts is, therefore, not an important channel in explaining the generational differences in teacher quality for the 1970-1987 cohorts.

A separate question from whether there is *changing* selective attrition across cohorts is whether constant selective attrition leads to the appearance of generational change where none exists. The structure of the data means that teachers in some earlier cohorts do not make it to our data unless they persist in teaching sufficiently. If teachers who persist have different permanent value-added than teachers who attrit, this will generate compositional differences between early and late cohorts.[33] It is worth noting that the results from Figure 2a showing differential improvement for teaching Black students among cohorts of teachers born between 1975 and 1992 *cannot* be driven by sample selection because these cohorts are fully observed in our data.[34]

Table 4 assesses the role of early- and mid-career attrition in the generational differences in teacher quality by comparing teachers who are more similar in their eventual persistence. For example, to address the concern that only some cohorts need to persist to 5 years to be included in our data, we can simply drop *all* teachers who fail to persist to 5 years and examine whether

[31] Interestingly, though the rate of attrition for high and low VA teachers across these cohorts is similar, in results not shown, we see that attrition rates for teachers who attended high or medium selectivity colleges are significantly higher (3.4 pp) than for teachers attending lower selectivity colleges.
[32] Among the Gen X teachers born between 1970-1979, 29% of low value-added teachers leave compared to 30% of high value-added teachers, but among Millennial teachers born between 1980-1987, 23% of low-value-added teachers and 22% of high value-added teachers leave, respectively.
[33] Past studies find mixed evidence on teacher value-added and attrition, with some suggesting that the best teachers are more likely to stay, e.g., Goldhaber, Gross, and Player (2011), and others like West and Chingos (2009) finding that higher value-added teachers leave the public school system at the same rate as less effective ones.
[34] The 1975 birth cohort turns 22 in 1997, the first year of our data, which is typically how old individuals entering teaching right after graduating from college would be.

the generational gaps in teacher quality change. Columns 1 and 2 of Table 4 show the results of this exercise, where Column 1 replicates our earlier analysis and Column 2 drops all teachers who fail to persist to 5 years.[35] The broad takeaway is similar in both columns suggesting that early-career selective attrition does not drive our estimates. That said, even in Column 2, it remains the case that there is scope for selective attrition to influence results as some cohorts require persistence to more than 5 years to be included in this sample, whereas other cohorts only require persistence to 5 years. To assess the role of mid-career attrition, in Columns 3 and 4, we impose the restriction that teachers persist to 8 or 10 years, respectively. Compared to Column 1, the gap between Baby Boomers and Gen X decreases slightly and the gap between Gen X and Millennials increases slightly, but on the whole, the qualitative story is similar suggesting that mid-career attrition does not drive our findings. Though Column 4 does not rule out late-career sample selection entirely, it fully addresses sample selection for the Millennial versus Gen X comparison. Specifically, teachers are included in Column 4 only if they persist to 10 years, and none of the Gen X cohorts face a more stringent persistence requirement (beyond 10 years) to be observed in our sample.[36]

That said, many Baby Boomer cohorts still face different persistence requirements in order to be observed in the sample in Column 4, so this analysis does not fully rule out the potential effects of selective attrition on the estimated gaps between Baby Boomers and recent cohorts. However, the selective attrition would have to be differential by effectiveness for teaching students of different race and concentrated entirely in the late career period.

[35] Unlike the analysis in Appendix Table 2 that dropped teacher-years with low levels of experience, here we make restrictions at the teacher level and include all teacher-years for any teacher who persists sufficiently.

[36] The earliest Gen X cohort (born 1965) turns 22 in 1987 and therefore is included in our data in 1997 as long as they persist for 10 years. Requiring every cohort to persist to 10 years thus imposes the same persistence requirement on Millennials and Gen X.

6.2 The role of teacher characteristics

We assess whether observable teacher characteristics can explain the generational differences in quality. Appendix Table A4 adds controls for teacher gender, race/ethnicity, whether the teacher ever has an advanced degree, indicators for selectivity of college attended, and whether the teacher obtained a BA from an institution in the South, as well as interactions of each of these controls with R_b, the indicator that the value-added measure is for black students to Equation (3). Consistent with past findings that observed teacher characteristics explain little of the variation in teacher quality, we find that these characteristics do not mediate the generational differences in quality.

We also investigate heterogeneity in the generational differences in teacher quality across demographic groups, which may shed light on potential mechanisms. For example, school segregation changed more in the South than in other parts of the country, so teachers originally from the South may drive the effect to the extent that this is a central mechanism. We also assess generational differences in teacher quality for white and Black teachers separately. Compositional changes are likely to have been large for Black teachers across these generations since the fraction of teachers who are Black dropped from 18% among Baby Boomers to 8% among Millennials. There were also significant improvements to both the quality of K-12 schools (as a result of desegregation) and colleges attended by Black teachers.[37] Black and white teachers may also exhibit different patterns of generational change for teaching black students because changes in racial attitudes and the adoption of CRP are likely to be especially relevant for white teachers.

[37] Though we find large differences in the quality of undergraduate institutions attended by Black teachers, past studies find that degree granting institution is not strongly correlated with teacher value-added so this may not result in differences in teacher quality on its own unless the shifts in institution attended are correlated with other changes.

Table 5 shows generational differences in teacher quality separately for each of these demographic groups. We use an indicator for obtaining a BA from a southern institution as a crude proxy for whether the teacher grew up in the south.[38] The qualitative pattern of generational differences in teacher quality is present in all four subgroups. The first two columns show that the effects seen for teacher quality for teaching Black students are fairly similar for teachers with a BA from the south as for teachers who got their BA outside the south, leaving open the possibility that common changes experienced by cohorts nationwide play a role.

Turning to the effects by teacher's race, the finding of differential improvement in teaching Black students versus white students is driven mostly by white teachers. Perhaps unsurprisingly, we see that this differential race-specific improvement is more muted among Black teachers, particularly for Millennials. The main effects for Black teachers indicate that recent generations of Black teachers are substantially more effective than Baby Boomer Black teachers, which could be due to either the improvements in education for Blacks and/or changes in who goes into teaching across the generations.

6.3 Teacher biases

Lastly, we investigate generational differences in teacher biases. First, we implement a test developed by Rangel and Shi (2021) to detect racial evaluation bias and assess whether it varies across teacher generation. Studies show that teacher expectations impact student achievement and academic trajectories (Hill and Jones, 2021; Lavy and Sand, 2018; Papageorge et al., 2020). For the years 2006-2013, the data include teacher evaluations of how they predict students will perform on their end-of-grade (EOG) tests which we use in conjunction with the students' objective, blindly scored performance to identify each teacher's evaluation bias for her Black

[38] Naturally, some teachers originally from other regions may attend college in the south and teachers originally from the south may attend college in a different region but this is the only proxy in the data.

students versus her white students. Teachers rate students on a non-interval scale with categories of insufficient mastery, inconsistent mastery, consistent mastery or superior performance. As discussed in and empirically assessed in Rangel and Shi (2021), the teacher evaluations and the blind scores should measure the same underlying skills.[39]

We estimate racial evaluation bias with the following regression:

$$E_{ijt} = \alpha_1 Black_i + \alpha_2 Black_i * Gen\,X_j + \alpha_3 Black_i * Mill_j + f(EOG)_{it} + X'_{ijt}\beta + \tau_{jt} + \varepsilon_{ijt} \quad (5)$$

where E_{ijt} is one of three indicators capturing teacher *j*'s evaluation of student *i*'s math achievement in year *t*. These indicators are whether predicted performance is above consistent, below consistent or well below consistent.[40] The regression includes white and Black students, and controls for student *i*'s raw math EOG test score in year *t*, X_{ijt} are student covariates including gender, indicators for special education and limited English proficiency status, and indicators for the student's relative age to the modal age in the grade, and τ_{jt} are teacher-by-year fixed effects.[41] Our primary coefficients of interest are α_1, α_2, and α_3. The coefficient α_1 estimates the average evaluation gap for Black students versus white students with the *same* EOG score (and other controlled characteristics) among Baby Boomer teachers, the omitted generation category. The coefficients, α_2 and α_3, estimate the difference in this Black-white evaluation gap for Gen X teachers and Millennial teachers versus Baby Boomer teachers, respectively.

---

[39] The tests aim to measure student proficiency at each grade level. The tests are scored electronically, and the raw test scores then mapped to achievement categories of insufficient mastery, inconsistent mastery, consistent mastery, and superior performance. On the day of testing, teachers are asked to predict each student's achievement category on the exam.
[40] "Above consistent" corresponds to superior performance. "Below consistent" corresponds to either inconsistent or insufficient mastery. "Well below consistent" corresponds to insufficient mastery.
[41] We control for EOG scores flexibly using grade-by-year-by-raw score fixed effects. Since we find generational differences in teacher quality for math only, we present results for teachers' racial evaluation bias in math.

Columns 1-3 of Table 6 present estimates of Equation (5). The coefficients on the indicator for whether the student is Black in these columns show that Baby Boomer teachers rate Black students lower relative to white classmates. These teachers are 2.4 percentage points more likely to evaluate a Black student as being below consistent, and 2.9 percentage points less likely to evaluate a Black student as being above consistent compared to a white student with the same test scores, respectively.

Our main interest lies in whether there are generational differences in this racial evaluation bias. There are no significant generational differences in teachers' racial evaluation bias when it comes to expecting that a student is below consistent or well below consistent. However, compared to Baby Boomer teachers, both Gen X and Millennial teachers have significantly lower racial evaluation bias when expecting above consistent performance. The coefficient on the interaction of Black student with Millennial teacher in Column 3 is 1.2 percentage points, which suggests that Millennial teachers exhibit approximately 40% less racial evaluation bias than that seen among Baby Boomer teachers. The generational difference in racial evaluation bias is seen only on one margin and the effects are too small to explain the generational differences in teacher quality on their own.[42] That said, we view the generational difference in racial evaluation bias as suggestive of other tangible differences across teachers from different cohorts, that likely affect student outcomes.

We supplement our analysis of racial evaluation bias for the teachers in our sample with data from the Project Implicit Database to describe birth cohort differences in implicit racial bias among US teachers more generally.[43] The implicit association test, IAT, is a computer test that

[42] Hill and Jones (2021) show that having high expectations increases test scores but scaling their estimates by our 0.012 effect on having high expectations leads to a trivially small effect.
[43] An important caveat for the IAT analysis is that the Project Implicit Database is not a representative sample because individuals self-select into taking the test.

measures implicit racial bias by using the difference in reaction times when individuals are asked to associate positive and negative attributes with pictures of Black or white faces. Figure 5 shows birth cohort differences in the fraction of teachers and non-teachers who show moderate to severe bias on the race IAT, i.e., an IAT score greater than 0.35.[44] Both teachers and non-teachers show clear differences across birth cohorts with approximately 51% of the Baby Boomer teachers showing moderate to severe implicit bias, but only 41% of Millennial cohorts exhibiting this amount of implicit bias.

Recent studies find that implicit bias affects student achievement. For example, Carlana (2019) finds that girls significantly underperform in math when assigned to teachers with stronger gender implicit bias, and Chin et al. (2020) find that implicit racial bias is associated with larger Black-white test score gaps. As such, the generational differences in teachers' implicit racial bias are a conceptually plausible mechanism that could contribute to the generational differences in teacher quality we find.

7. Conclusion

This study finds that recent cohorts of teachers are significantly more effective than earlier cohorts of teachers. Furthermore, the improvement in quality for teaching Black students across the cohorts is roughly twice that seen for teaching white students. The constellation of evidence across our empirical assessments is consistent with differences in teachers' racial attitudes and biases contributing to the race-specific improvements in teacher quality. Future research should further investigate potential determinants for these differences in teacher quality, particularly focusing on understanding whether these differences in quality are due to differences in teacher behavior that are easily changed or due to more static differences. Qualitative work describing

[44] We follow the literature in how it typically defines bias thresholds (Greenwald et al. 2003, Alesina et al. 2018).

differences in teaching practices or attitudes towards race across teacher cohorts may be particularly promising.

Understanding differences in teacher quality across cohorts can help policymakers understand how teacher quality is likely to evolve as current teachers retire. There is reason to be optimistic about the evolution of teacher quality in the coming years since the impending retirements of the earlier cohorts of teachers should naturally lead to improved student achievement and a narrowing of the Black-white test score gap.

Bibliography

Aizer, Anna, Ryan Boone, Adriana Lleras-Muney, and Jonathan Vogel. *Discrimination and racial disparities in labor market outcomes: Evidence from WWII*. No. w27689. National Bureau of Economic Research, 2020.

Alesina, A., Carlana, M., La Ferrara, E. and Pinotti, P., 2018. *Revealing stereotypes: Evidence from immigrants in schools* (No. w25333). National Bureau of Economic Research.

Allen, V. S. (1971). An analysis of textbooks relative to the treatment of Black Americans. *The Journal of Negro Education*, *40*(2), 140-145.

Allport, Gordon W. (1954), The Nature of Prejudice, Addison-Wesley, Oxford, England. Pages: xviii, 537.

Aneja, Abhay, and Guo Xu. "The Costs of Employment Segregation: Evidence from the Federal Government Under Woodrow Wilson." *The Quarterly Journal of Economics* 137, no. 2 (2022): 911-958.

Ashenfelter, O., Collins, W.J. and Yoon, A., 2006. Evaluating the role of Brown v. Board of Education in school equalization, desegregation, and the income of African Americans. *American Law and Economics Review*, *8*(2), pp.213-248.

Bacher-Hicks, A., Kane, T.J. and Staiger, D.O., 2014. *Validating teacher effect estimates using changes in teacher assignments in Los Angeles* (No. w20657). National Bureau of Economic Research.

Bacher-Hicks, A., Chin, M.J., Kane, T.J. and Staiger, D.O., 2019. An experimental evaluation of three teacher quality measures: Value-added, classroom observations, and student surveys. *Economics of Education Review*, *73*, p.101919.

Bacolod, M.P., 2007. Do alternative opportunities matter? The role of female labor markets in the decline of teacher quality. *The Review of Economics and Statistics*, *89*(4), pp.737-751.

Black, S.E. and Juhn, C., 2000. The rise of female professionals: Are women responding to skill demand? *American Economic Review*, *90*(2), pp.450-455.

Bobo, L.D., Charles, C.Z., Krysan, M. and Simmons, A.D., 2012. 3. The Real Record on Racial Attitudes. In *Social trends in American life* (pp. 38-83). Princeton University Press.

Boisjoly, J., Duncan, G.J., Kremer, M., Levy, D.M. and Eccles, J., 2006. Empathy or antipathy? The impact of diversity. *American Economic Review*, *96*(5), pp.1890-1905.
Borjas, George J., 1987. *Self-selection and the earnings of immigrants*. No. w2248. National Bureau of Economic Research.

Burns, J., Corno, L. and La Ferrara, E., 2015. Interaction, prejudice, and performance. Evidence from South Africa. *Working Paper*.

Carlana, M., 2019. Implicit stereotypes: Evidence from teachers' gender bias. *The Quarterly Journal of Economics*, *134*(3), pp.1163-1224.

Cascio, E., Gordon, N., Lewis, E. and Reber, S., 2008. From Brown to busing. *Journal of Urban Economics*, *64*(2), pp.296-325.

Cascio, E., Gordon, N., Lewis, E. and Reber, S., 2010. Paying for progress: Conditional grants and the desegregation of southern schools. *The Quarterly Journal of Economics*, *125*(1), pp.445-482.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, *104*(9), 2633-79.

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

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, *26*(6), 673-682.

Cook, Lisa D. "Violence and economic activity: evidence from African American patents, 1870–1940." *Journal of Economic Growth* 19, no. 2 (2014): 221-257.

Corcoran, S. P., Evans, W. N., & Schwab, R. M. (2004). Women, the labor market, and the declining relative quality of teachers. *Journal of Policy Analysis and Management*, *23*(3), 449-470.

Darling-Hammond, L., Bastian, K., Berry, B., Carver-Thomas, D., Kini, T., Levin, S., & McDiarmid, W. (2019). Educator Supply, Demand, and Quality in North Carolina: Current Status and Recommendations. Palo Alto, CA: Learning Policy Institute.

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

Dinerstein, M., Megalokonomou, R. and Yannelis, C., 2022. Human capital depreciation and returns to experience. *American Economic Review*, *112*(11), pp.3725-62.

Edmonds, Lavar. (2022). Role Models Revisited: HBCUs, Same-Race Teacher Effects, and Black Student Achievement.

Gitomer, D. H. (2007). Teacher quality in a changing policy landscape: Improvements in the teacher pool (ETS Policy Information Report). Princeton, NJ: Educational Testing Service. Retrieved from https://www.ets.org/Media/Research/pdf/PICTQ.pdf

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

Goldhaber, D., & Hansen, M. (2010). Race, gender, and teacher testing: How informative a tool is teacher licensure testing?. *American Educational Research Journal*, 47(1), 218-251.

Goldhaber, D., & Walch, J. (2013). Rhetoric versus reality: Is the academic caliber of the teacher workforce changing? *Center for Education Data and Research, University of Washington, Bothell*, 202013-4.

Goldin, C., Understanding the Gender Gap: An Economic History of American Women (New York, NY: Oxford University Press, 1990). "Career and Family: College Women Look to the Past," in Francine Blau and Ronald Ehrenberg (Eds.), Gender and Family Issue

Greenwald, A.G., Poehlman, T.A., Uhlmann, E.L. and Banaji, M.R., 2009. Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of personality and social psychology*, *97*(1), p.17.

Guryan, J., 2004. Desegregation and Black dropout rates. *American Economic Review*, *94*(4), pp.919-943.

Hanushek, E.A., Kain, J.F. and Rivkin, S.G., 2009. New evidence about Brown v. Board of Education: The complex effects of school racial composition on achievement. *Journal of Labor Economics*, *27*(3), pp.349-383.

Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, *95*(7-8), 798-812.

Hill, A.J. and Jones, D.B., 2021. Self-fulfilling prophecies in the classroom. *Journal of Human Capital*, *15*(3), pp.400-431.

Hinrichs, P., 2016. An empirical analysis of racial segregation in higher education (No. w21831). *National Bureau of Economic Research*.

Hinrichs, P., 2021. What kind of teachers are schools looking for? Evidence from a randomized field experiment. *Journal of Economic Behavior & Organization*, *186*, pp.395-411.

Hoxby, C.M. and Leigh, A., 2004. Pulled away or pushed out? Explaining the decline of teacher aptitude in the United States. *American Economic Review Papers and Proceedings*, *94*(2), pp.236-240.

Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow. "The allocation of talent and us economic growth." *Econometrica* 87, no. 5 (2019): 1439-1474.

Jacob, B.A., Rockoff, J.E., Taylor, E.S., Lindy, B. and Rosen, R., 2018. Teacher applicant hiring and teacher performance: Evidence from DC public schools. *Journal of Public Economics*, *166*, pp.81-97.

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

Jackson, C.K., Rockoff, J.E. & Staiger, D.O., 2014. Teacher effects and teacher-related policies. *Annual Review of Economics*, *6*(1), pp.801-825.

Jackson, C.K., 2018. What do test scores miss? The importance of teacher effects on non–test score outcomes. *Journal of Political Economy*, *126*(5), pp.2072-2107.

Johnson, R.C., 2011. Long-run impacts of school desegregation & school quality on adult attainments (No. w16664). *National Bureau of Economic Research*.

Kane, T. J., & Staiger, D. O. (2008). Estimating teacher impacts on student achievement: An experimental evaluation (No. w14607). *National Bureau of Economic Research*.

Kane, T.J., McCaffrey, D.F., Miller, T. and Staiger, D.O., 2013. Have we identified effective teachers? Validating measures of effective teaching using random assignment. In *Research Paper. MET Project. Bill & Melinda Gates Foundation*.

Kilpatrick, J. (1992). A history of research in mathematics education. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 3–38). New York: Macmillan

Lankford, H., Loeb, S., McEachin, A., Miller, L. C., & Wyckoff, J. (2014). Who enters teaching? Encouraging evidence that the status of teaching is improving. *Educational Researcher*, *43*(9), 444-453.

Lavy, V. and Sand, E., 2018. On the origins of gender gaps in human capital: Short-and long-term consequences of teachers' biases. *Journal of Public Economics*, *167*, pp.263-279.

McCaffrey, D. F., Lockwood, J. R., Mihaly, K., & Sass, T. R. (2012). A review of Stata commands for fixed-effects estimation in normal linear models. *The Stata Journal*, *12*(3), 406-432.

Milner, H. R. (2011). Culturally relevant pedagogy in a diverse urban classroom. *The Urban Review*, *43*(1), 66-89.

Nagler, M., Piopiunik, M. and West, M.R., 2020. Weak Markets, Strong teachers: Recession at Career Start and Teacher Effectiveness. *Journal of Labor Economics*, *38*(2), pp.453-500.

Ost, B., 2014. How do teachers improve? The relative importance of specific and general human capital. *American Economic Journal: Applied Economics*, *6*(2), pp.127-51.

Papay, J.P. and 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*, pp.105-119.
Papageorge, N.W., Gershenson, S. and Kang, K.M., 2020. Teacher expectations matter. *Review of Economics and Statistics*, *102*(2), pp.234-251.

Rangel, M.A. and Shi, Y., 2021. First Impressions: The Case of Teacher Racial Bias.

Reber, S. J., 2010. School desegregation and educational attainment for Blacks. *Journal of Human resources*, *45*(4), pp.893-914.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417-458.

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

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The urban review*, *3*(1), 16-20.

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics*, *125*(1), 175-214.

Rothstein, J. (2017). Measuring the impacts of teachers: comment. *American Economic Review*, *107*(6), 1656-84.

Sidanius, J., Levin, S., Van Laar, C. and Sears, D.O., 2008. *The diversity challenge: Social identity and intergroup relations on the college campus*. Russell Sage Foundation.

Taylor, P., Funk, C. & Craighill P. (2006). Guess Who's Coming to Dinner. 22% of Americans have a Relative in a Mixed-Race Marriage. *Pew Research Center Social Trends Report.*

Temin, Peter, "Teacher Quality and the Future of America*," Eastern Economics Journal* 28 (2002), 285–300.

Thompson, R. L. (1979). *A Historical And Legal Analysis Of Teacher Certification In North Carolina*. The University of North Carolina at Greensboro.

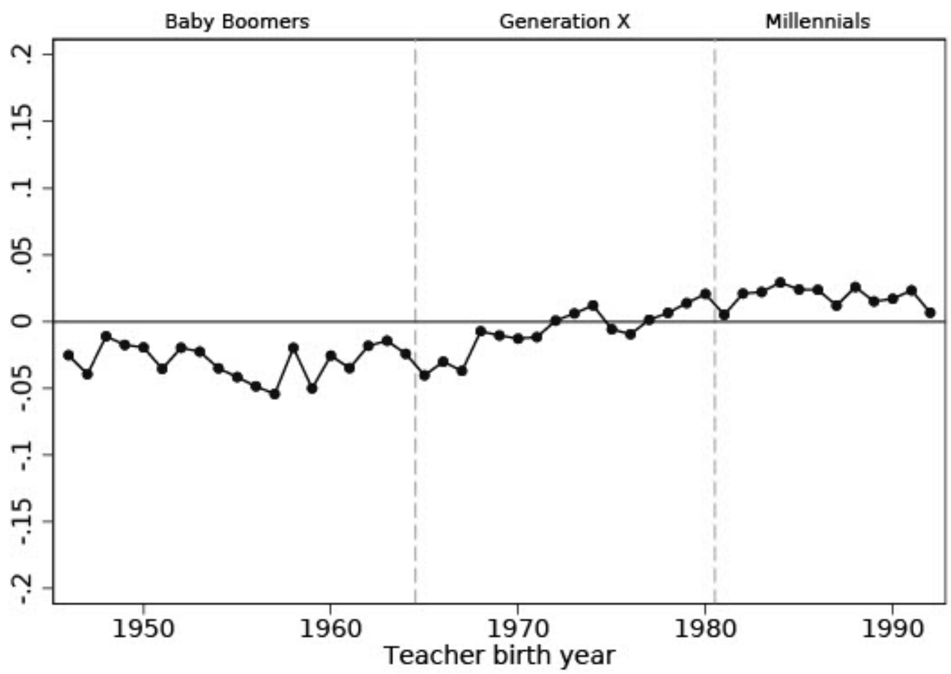
Van Laar, C., Levin, S., Sinclair, S. and Sidanius, J., 2005. The effect of university roommate contact on ethnic attitudes and behavior. *Journal of Experimental Social Psychology*, *41*(4), pp.329-345.

Weiner, D.A., Lutz, B.F. and Ludwig, J., 2009. The effects of school desegregation on crime (No. w15380). *National Bureau of Economic Research*.

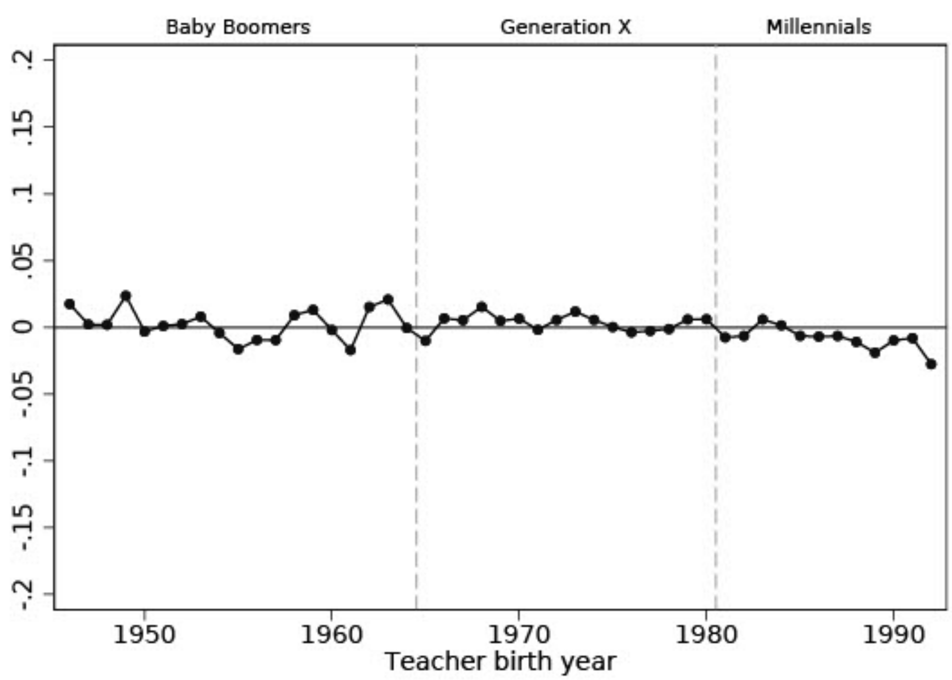
West, Martin R., and Matthew M. Chingos. "Teacher effectiveness, mobility, and attrition in Florida." *Performance Incentives: Their Growing Impact on American K-12 Education* (2009): 251-71.

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

Figure 1: Teacher Value-Added and Teacher Birth Year



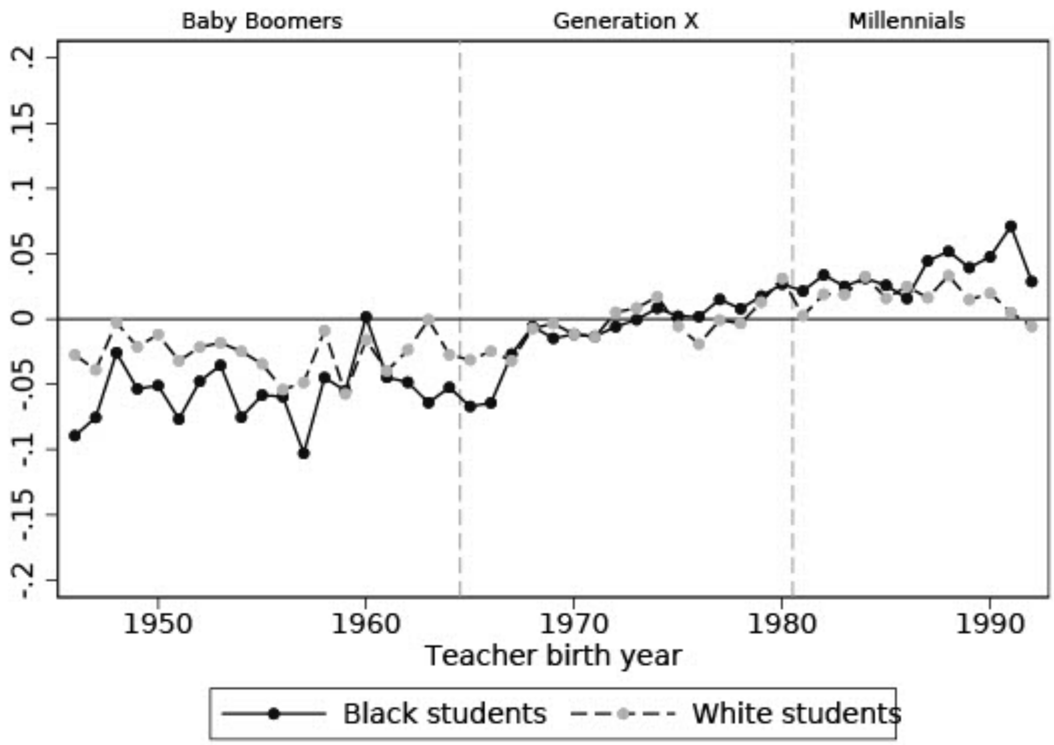
Panel a) Math value-added



Panel b) Reading value-added

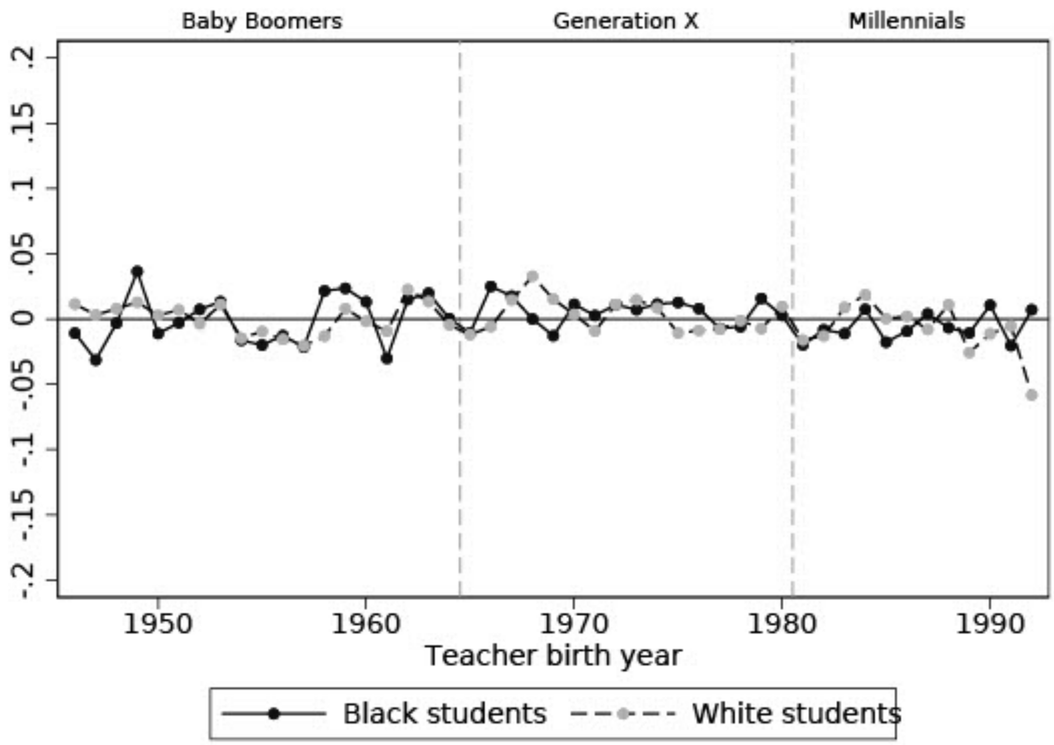
Notes: Average value-added in math (panel a) and reading (panel b) by teacher birth year cohort. Estimates of value-added from Equation (1) which is estimated for all students. Value-added estimates are normalized to have mean zero.

Figure 2a: Race-Specific Teacher Value-added in Math and Teacher Birth Year



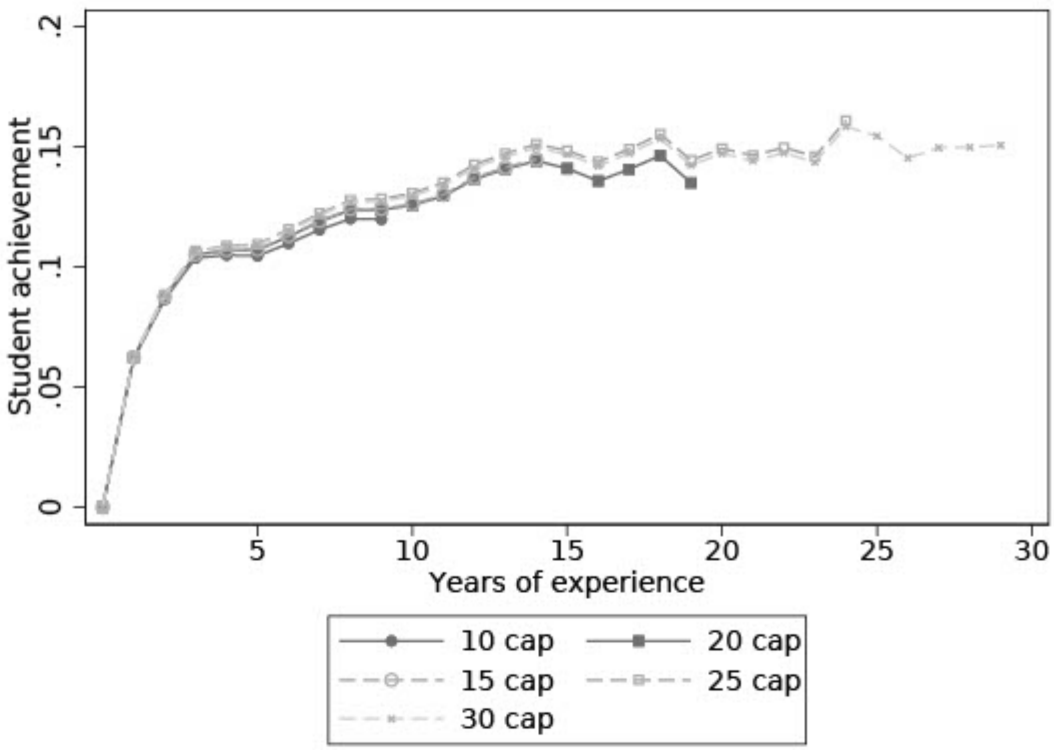
Notes: Average race-specific math value-added by teacher birth year cohort. Estimates of value-added from Equation (1) estimated separately for Black students and white students. The value-added estimates for Black and white students are each normalized to have mean zero.

Figure 2b: Race-Specific Teacher Value-added in Reading and Teacher Birth Year



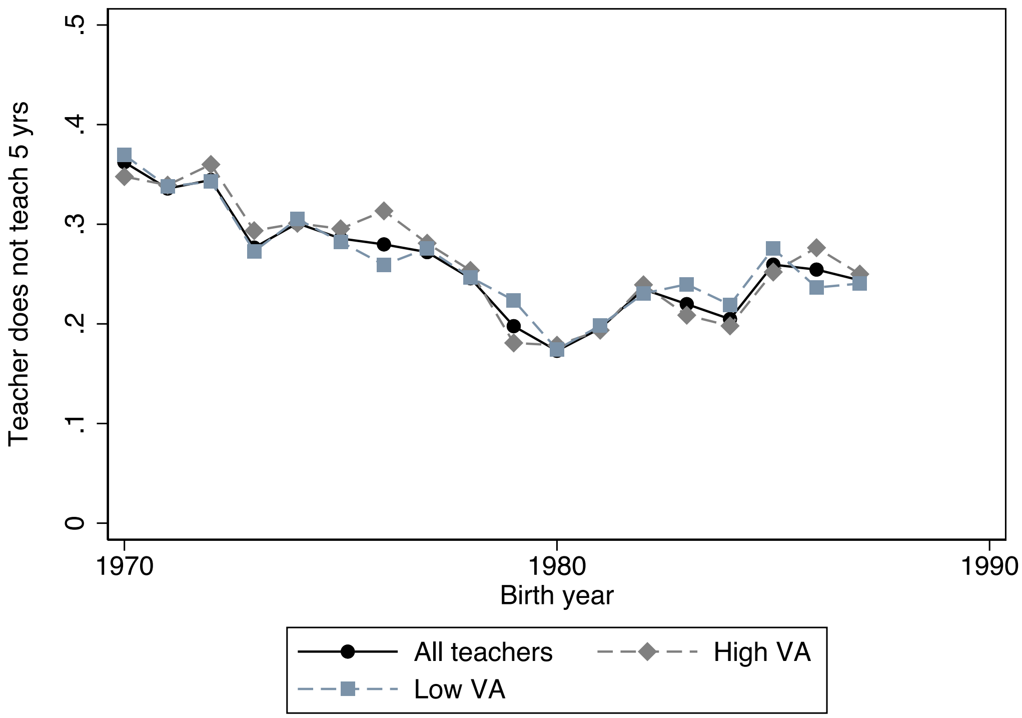
Notes: Average race-specific reading value-added by teacher birth year cohort. Estimates of value-added from Equation (1) estimated separately for Black students and white students. The value-added estimates for Black and white students are each normalized to have mean zero.

Figure 3: Experience Profiles for Student Achievement in Math using Various Caps to Experience Returns



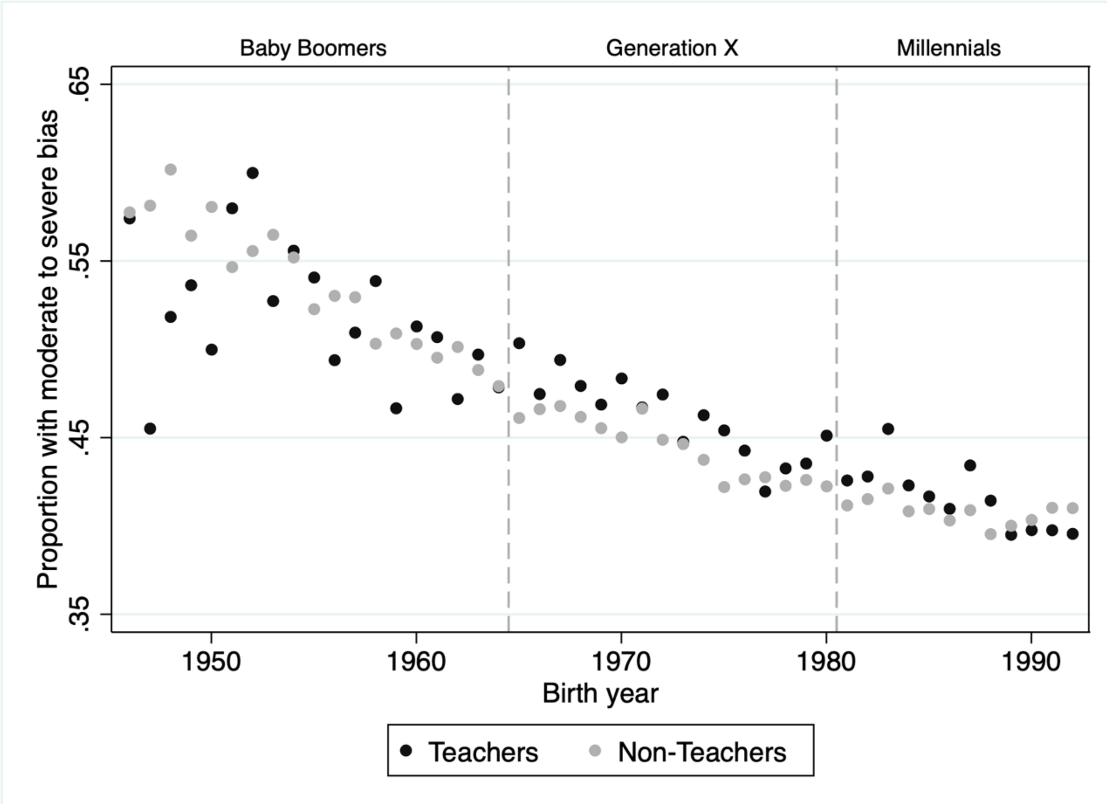
Notes: This figure reports coefficients on indicators for years of experience from the estimation of Equation (1). We show results from estimating five different specifications which alternate the cap imposed on experience returns at 10, 15, 20, 25 and 30 years.

Figure 4: Attrition by Teacher Birth Year and Teacher Value-Added



Notes: This figure shows the probability that the teacher fails to accumulate five years of experience by birth cohort (birth years 1970-1987) by teacher math value-added. High VA refers to teachers with math value-added above the median and low VA refers to teachers with math value-added below the median. We restrict attention to teachers who start as novices between 1997 and 2010, and measure whether a teacher accumulates 5 years of experience within 7 years of starting.

Figure 5: Differences in Implicit Racial Bias across Birth Cohorts



Notes: Author calculations from the Race Implicit Association Test (IAT) of the Project Implicit Database compiled by Xu, Nosek, and Greenwald (2014). The IAT is an online test: participation in the IAT is typically voluntary but some workplaces and schools require it. A higher race IAT score indicates higher implicit racial bias as it captures a stronger association between Black and negative attributes. The figure shows the proportion of teachers and non-teachers in the US with moderate to severe bias across different birth cohorts, 1946-1992. Moderate to severe bias is defined in the literature as having an IAT score higher than 0.35.

Table 1: Teacher Characteristics by Generation

	Baby Boomers	Generation X	Millennials
Definition: Birth year cohorts	1946-1964	1965-1980	1981-1992
Female teacher	0.94	0.90	0.91
Black teacher	0.18	0.11	0.08
Hispanic teacher	0.00	0.01	0.01
Other race teacher	0.01	0.02	0.03
Teacher has > Bachelor's	0.44	0.30	0.28
Experience	23.35	6.67	2.57
BA institution selectivity is low	0.28	0.18	0.12
BA institution selectivity is medium	0.45	0.50	0.54
BA institution selectivity is high	0.27	0.31	0.34
BA from Southern institution	0.93	0.73	0.73
Number of teachers	3,976	10,323	6,985

Notes: Summary statistics for the teachers in the analysis sample contributing to the estimation of Equation (1). The data includes teachers' demographic characteristics, identifiers for undergraduate institution attended, and yearly information on National Board Certification status, degrees earned, and years of experience. This table reports statistics on whether a teacher is ever observed as National Board certified or with a degree greater than a Bachelor's in the data. The years of experience reported is an average over the period a teacher is observed in the data. Data on undergraduate institution selectivity comes from IPEDS data on 2010 Carnegie classification summarizing institutions' undergraduate admissions selectivity between 2008 to 2010. This data allows us to categorize institutions into four measures of selectivity: low selectivity, medium selectivity, high selectivity, and unknown selectivity (not shown). Institutions with low selectivity are inclusive with test scores indicating that they extend educational opportunity to students with a wide range of academic preparation. Institutions with medium selectivity have first year students' test scores that place them in the middle two-fifths of baccalaureate institutions, and institutions with high selectivity have first-year student test scores which place them in the top fifth of baccalaureate institutions.

Table 2: Generational Differences in Teacher Value-Added

	Math (1)	Math (2)	Math (3)	Reading (4)	Reading (5)	Reading (6)
Gen X	0.027***	0.032***	0.027***	0.001	0.001	0.002
	(0.004)	(0.004)	(0.005)	(0.003)	(0.003)	(0.004)
Millennials	0.049***	0.056***	0.046***	-0.009**	-0.004	-0.004
	(0.005)	(0.005)	(0.006)	(0.003)	(0.004)	(0.005)
Gen X * VA for black students			0.030***			0.005
			(0.005)			(0.006)
Millennials * VA for black students			0.044***			0.001
			(0.006)			(0.007)
Sample	All students	Black and white students	Black and white students	All students	Black and white students	Black and white students
Number of observations	21,284	18,460	36,920	21,508	18, 680	37,360

Notes: *p<0.1, **p<0.05, ***p<0.01. This table displays generational differences in overall teacher value-added from equation (2) estimated using all students in columns (1) and (4). Columns (2) and (5) show generational differences in teacher value-added from equation (2) estimated using black and white students for teachers who have a valid race-specific value-added estimate for white students as well as black students. The number of observations in columns (1), (2), (4) and (5) corresponds to the number of teachers. Columns (3) and (6) present the estimates of generational differences in teacher value-added from equation (3). In columns (3) and (6), each teacher has two value-added estimates, one for her black students and one for her white students: the unit of observation is teacher-by-value added estimate and standard errors are clustered at the teacher level. Throughout this table, standard errors are shown in parentheses. The omitted group is Baby Boomer teachers in each specification.

Table 3: Generation Differences in Teacher Value-Added Within Schools

	(1)	(2)	(3)
Gen X	0.027***	0.023***	0.010*
	(0.005)	(0.005)	(0.005)
Millennials	0.046***	0.042***	0.030***
	(0.006)	(0.006)	(0.006)
Gen X * VA for Black students	0.030***	0.044***	0.042***
	(0.005)	(0.005)	(0.005)
Millennials * VA for Black students	0.044***	0.066***	0.066***
	(0.006)	(0.006)	(0.006)
Number of observations	36,920	46,142	46,142
Estimating teacher-by-school fixed effects	No	Yes	Yes
Within school comparison only	No	No	Yes

Notes: *p<0.1, **p<0.05, ***p<0.01. Column (1) shows the preferred results from estimating Equation (3). Columns (2) and (3) show results from estimating Equation (4). Column (2) does not include race-by-school fixed effects, but Column (4) does. Standard errors are clustered at the teacher level and are shown in parentheses. The omitted group is Baby Boomer teachers.

Table 4: Generational Differences in Teacher Value-Added: the Role of Career Attrition

	(1)	(2)	(3)	(4)
Gen X	0.027***	0.026***	0.024***	0.021***
	(0.005)	(0.005)	(0.005)	(0.005)
Millennials	0.046***	0.055***	0.051***	0.049***
	(0.006)	(0.006)	(0.007)	(0.009)
Gen X * VA for Black students	0.030***	0.029***	0.030***	0.029***
	(0.005)	(0.005)	(0.005)	(0.006)
Millennials * VA for Black students	0.044***	0.037***	0.034***	0.038***
	(0.006)	(0.007)	(0.008)	(0.010)
Number of observations	36,920	29,854	24,390	21,066
Teachers who teach for at least:	No restriction	5 years	8 years	10 years

Note: *p<0.1, **p<0.05, ***p<0.01. Estimates of Equation (3) which restrict the analysis sample to teachers who will persist in teaching in NC public schools for at least 5, 8, and 10 years, respectively. Standard errors are clustered at the teacher level and are shown in parentheses. The omitted group is Baby Boomer teachers.

Table 5: Heterogeneity in Generational Differences in Teacher Value-Added by Teacher Demographics

	(1)	(2)	(3)	(4)
Gen X	0.015	0.025***	0.012**	0.052***
	(0.019)	(0.006)	(0.005)	(0.015)
Millennials	0.032*	0.046***	0.025***	0.124***
	(0.019)	(0.006)	(0.006)	(0.021)
Gen X * VA for Black students	0.043***	0.029***	0.035***	0.025*
	(0.016)	(0.006)	(0.006)	(0.013)
Millennials * VA for Black students	0.058***	0.044***	0.053***	0.022
	(0.017)	(0.007)	(0.007)	(0.019)
Number of observations	8,922	27,972	31,682	4,186
Demographic group of teacher:	BA not in South	BA in South	White teachers	Black Teachers

Note: *p<0.1, **p<0.05, ***p<0.01. Estimates of Equation (3) estimated separately for teachers belonging to different demographic groups: teachers who got their undergraduate degree from an institution not in the South, teachers who got their undergraduate degree from an institution in the South, white teachers and Black teachers. Standard errors are clustered at the teacher level and are shown in parentheses. The omitted group is Baby Boomer teachers.

Table 6: Generational Differences in Racial Evaluation Bias

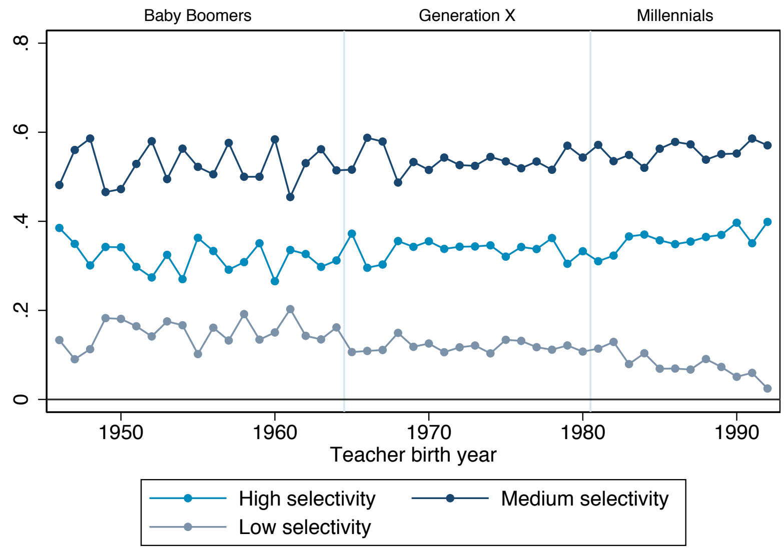
	Well below consistent (1)	Below consistent (2)	Above consistent (3)
Student is Black	0.007***	0.024***	-0.029***
	(0.002)	(0.004)	(0.003)
Gen X * Student is Black	-0.004	0.001	0.008**
	(0.002)	(0.004)	(0.004)
Millennials * Student is Black	-0.001	-0.002	0.012***
	(0.003)	(0.004)	(0.004)
Number of observations	638,630	638,630	638,630

Notes: *p<0.1, **p<0.05, ***p<0.01. This table displays the estimates of generational differences in racial evaluation bias from Equation (5). Teacher evaluation information is available from 2006-2013. We vary the dependent variable in Columns 1-3 from an indicator for whether the teacher's evaluation of student achievement is above consistent, below consistent, or well below consistent. "Above consistent" corresponds to superior performance. "Below consistent" corresponds to either inconsistent or insufficient mastery. "Well below consistent" corresponds to insufficient mastery. All regressions control for the student's math EOG test score (grade-by-year-by-raw score fixed effects), student covariates including gender, indicators for special education and limited English proficiency status as well as indicators for the student's relative age to the modal age in the grade, and teacher-by-year fixed effects. Standard errors are clustered at the teacher level and are shown in parentheses. The omitted group is Baby Boomer teachers in each specification.

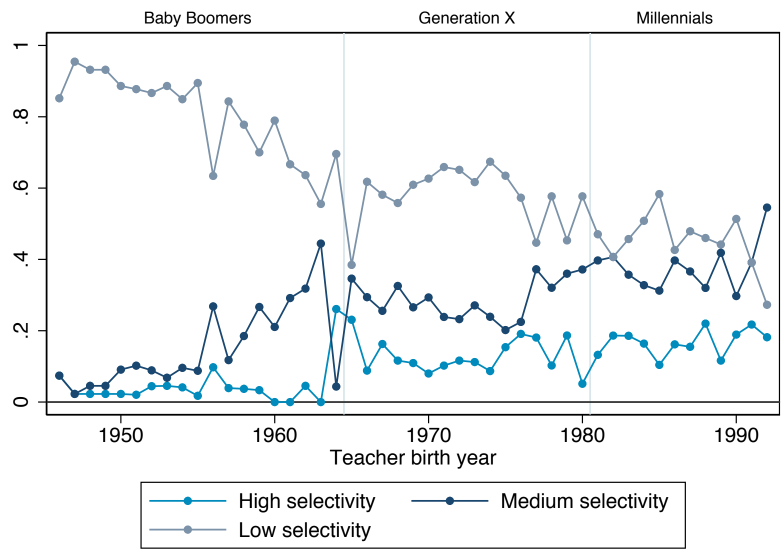
For Online Publication

Appendix Figure A1: Undergraduate Institution Selectivity and Teacher Birth Cohort

Panel A: White teachers

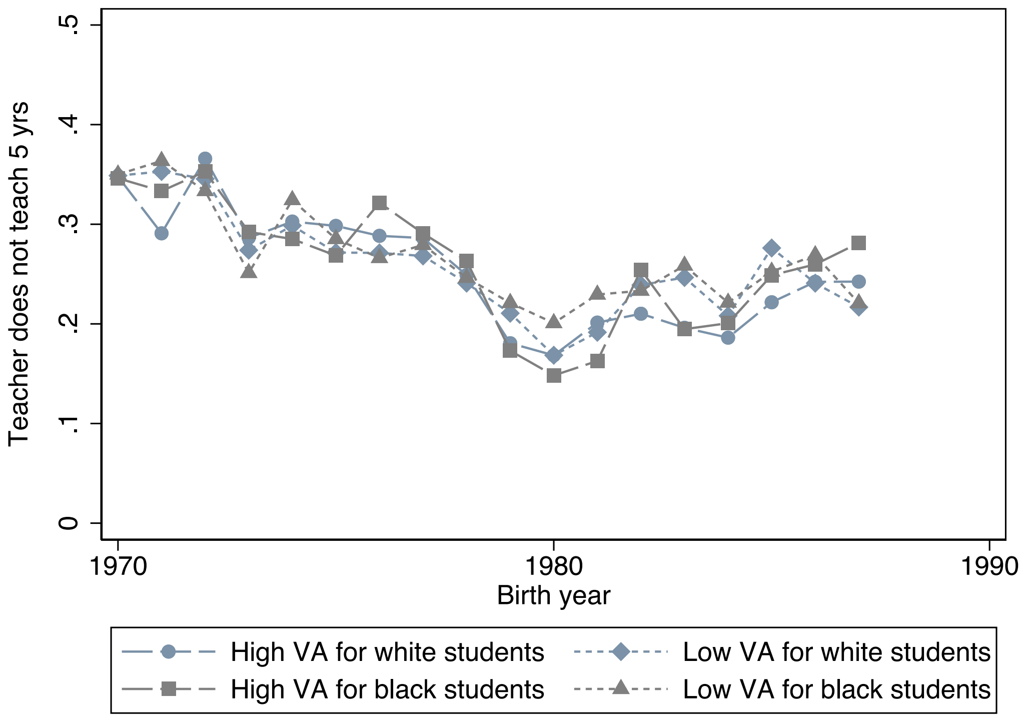


Panel B: Black teachers



Notes: Fraction of white teachers (Panel A) and Black teachers (Panel B) attending undergraduate institutions of various selectivity by teacher birth year cohort. Institution selectivity comes from the IPEDS 2010 Carnegie classification summarizing institutions' undergraduate admissions selectivity between 2008 to 2010. This data allows us to categorize institutions as follows: low selectivity, medium selectivity, high selectivity, and unknown selectivity (not shown). Institutions with low selectivity are inclusive, with test scores indicating that they extend educational opportunity to students with a wide range of academic preparation. Institutions with medium selectivity have first year students' test scores that place them in the middle two-fifths of baccalaureate institutions, and institutions with high selectivity have first-year student test scores which place them in the top fifth of baccalaureate institutions.

Appendix Figure A2: Attrition by Teacher Birth Year and Race-Specific Teacher Value-Added



Notes: This figure shows the probability that the teacher fails to accumulate five years of experience by birth cohort (birth years 1970-1987) by race-specific teacher math value-added. High VA for white students (Black students) refers to teachers with math value-added above the median for white students (Black students) and low VA for white students (Black students) refers to teachers with math value-added below the median for white students (Black students). We restrict attention to teachers who start as novices between 1997 and 2010, and measure whether a teacher accumulates 5 years of experience within 7 years of starting.

Appendix Table A1: Summary Statistics of Student Characteristics in the Estimation Sample

	Mean	SD
Math score	0.09	0.96
Reading score	0.07	0.97
Lagged math score	0.10	0.95
Lagged reading score	0.09	0.95
Female	0.52	0.50
White	0.59	0.49
Black	0.26	0.44
Hispanic	0.08	0.26
Other race	0.06	0.24
Special education	0.03	0.17
Limited English proficiency	0.01	0.12
Observations	1,567,716	

Notes: Summary statistics for the student-year observations which contribute to the estimation of Equation (1). The analysis sample comprises of fourth and fifth grade students.

Appendix Table A2: Robustness of Generational Differences in Teacher Value-Added

	Varying experience caps				Removing Exp>30
Gen X	0.027***	0.028***	0.036***	0.034***	0.024***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Millennials	0.046***	0.048***	0.061***	0.057***	0.042***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Gen X * VA for black students	0.030***	0.040***	0.042***	0.040***	0.036***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Millennials * VA for black students	0.044***	0.060***	0.063***	0.059***	0.055***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Experience capped at year:	15	20	25	30	15
Number of observations	36,920	36,920	36,920	36,920	36,578

Notes: *p<0.1, **p<0.05, ***p<0.01. Estimates of Equation (3) from specifications which vary the placement of the cap on the returns to experience from 15, 20, 25 to 30 years. The last column includes estimates of Equation (3) using the preferred cap of 15 but limits the analysis to observations for teachers with 30 years of experience or less. Standard errors are clustered at the teacher level and are shown in parentheses. The omitted group is Baby Boomer teachers.

Appendix Table A3: Generational Differences in Teacher Value-Added Restricted to Teacher-Years with Sufficient Experience

	(1)	(2)	(3)	(4)	(5)
Gen X	0.027***	0.026***	0.029***	0.027***	0.028***
	(0.005)	(0.006)	(0.006)	(0.006)	(0.008)
Millennials	0.046***	0.065***	0.072***	0.062***	
	(0.006)	(0.008)	(0.013)	(0.020)	
Gen X * VA for Black students	0.030***	0.029***	0.018***	0.022***	0.029***
	(0.005)	(0.006)	(0.006)	(0.007)	(0.009)
Millennials * VA for Black students	0.044***	0.047***	0.041***	0.033	
	(0.006)	(0.009)	(0.014)	(0.024)	
Number of observations	36,920	22,734	17,256	14,566	9,954
Teachers with experience of at least:	-	5	8	10	15

Notes: *p<0.1, **p<0.05, ***p<0.01. Preferred estimates of Equation (3) presented in Column (1). Estimates of generational differences in teacher value-added from Equation (3) which limit the analysis to observations for teachers with varying experience ranging from at least 5, 8, 10 to 15 years presented in Columns (2) through (5). Each teacher has two value-added estimates, one for her Black students and one for her white students, and the unit of observation is teacher-by-value-added estimate. Standard errors are clustered at the teacher level and are shown in parentheses. The omitted group is Baby Boomer teachers in each specification.

Appendix Table A4: Teacher Value-Added and Generation: the Role of Observed Teacher Characteristics

	(1)	(2)
Gen X	0.023***	0.021***
	(0.005)	(0.005)
Millennials	0.041***	0.038***
	(0.006)	(0.006)
Gen X * VA for Black students	0.031***	0.035***
	(0.005)	(0.006)
Millennials * VA for Black students	0.044***	0.050***
	(0.006)	(0.006)
Number of observations	36,168	36,168
Controls for teacher characteristics:	_	Teacher gender, race, advanced degree, college selectivity, BA in South

Notes: *p<0.1, **p<0.05, ***p<0.01. This table assesses how estimates of generational differences in teacher value-added change when controlling for observed characteristics of teachers. Column (2) estimates the generational differences in teacher value-added from Equation (2) which additionally control for teacher characteristics and interactions between the teacher characteristics and R_b, the indicator that the value-added measure is for black students. The teacher characteristics include teacher's gender, race, whether they ever have an advanced degree, indicators for the selectivity of college attended, and whether they obtained their BA from an institute in the South. Column (1) presents estimates of the generational differences in teacher value-added from Equation (3) for the subsample of teachers with non-missing information on the characteristics included in Column (2) for comparison purposes. Standard errors are clustered at the teacher level and are shown in parentheses. The omitted group is Baby Boomer teachers.