Why Black Teachers Matter

Abstract

Black teachers are critical resources for our children and schools. Pairing experimental data with rich measures of teacher mindsets and practices and varied student outcomes, I document that: (1) Black teachers in upper-elementary grades have large effects on the self-efficacy and classroom engagement of their Black students (0.7 and 0.8 SD) but not for non-Black students, potentially driven by role modeling; (2) race-matching effects on Black students' social-emotional learning explain a moderate to large share of effects on more distal outcomes, including absences and test scores; (3) Black teachers also benefit the test scores (0.2 SD) and absences (over 20% decrease) of all students—no matter their race/ethnicity—that often persist many years later into high school; and (4) in addition to potential role-modeling channels, Black teachers bring unique mindsets and practices to their work (e.g., preparation for and differentiated instruction, growth mindset beliefs, well-organized classrooms) that mediate a moderate to large share of their effects on student outcomes. These findings help bridge the quantitative "teacher like me" literature with theoretical discussion and qualitative exploration on why Black teachers matter.

Keywords: Race/ethnicity matching, teacher quality, culturally responsive teaching, experiment

Introduction

For far too long, education systems have failed students of color. Systemic racism—exhibited through school-based segregation (Johnson, 2011), exclusionary discipline (Fenning & Rose, 2007), limited access to instructional resources (Jackson, 2009), among other sources—has created stark disparities in educational opportunity between Black and other historically marginalized and minoritized students of color versus their White peers. Constrained opportunity impacts and ripples across a range of educational and life outcomes, including academic performance (Fryer & Levitt, 2004), high school graduation (Hernandez, 2011), and success in the labor market (Rivkin, 1995).

Compelling lines of theoretical and empirical research show that one of the most effective levers to better support Black and other students of color is to provide opportunities to learn from a teacher from the same racial or ethnic group. While the teacher workforce is overwhelmingly White (roughly 80%; U.S. Department of Education, 2019), Black and other teachers of color are described as uniquely positioned to understand and address the social, political, and economic inequalities that students of color face (Ladson-Billings, 1994). Building from this theory, causally oriented studies document substantively meaningful teacher-student race/ethnicity-matching effects on students' academic outcomes (for reviews, see Bristol & Martin-Fernandez, 2019; Redding, 2019).

In the current analyses, I orient the quantitative teacher-student race/ethnicity-matching literature towards "why" and "how" questions, which is important for at least two reasons. First, the theoretical literature on this topic poses several, likely overlapping hypotheses related to social dynamics inside schools and classrooms that are largely untested quantitatively. Is it that Black teachers serve as role models for their Black students, which in turn drives improved outcomes (Villegas & Lucas, 2004)? Do Black teachers also engage in "culturally relevant" (Ladson-Billings, 1995b), "culturally responsive" (Gay, 2000), and "culturally sustaining" (Paris, 2012) pedagogies that are particularly beneficial for their students of color—or that may benefit all students?

Second, understanding which of these mechanisms—or others—drive effects of Black teachers on student outcomes is critical for policy and practice. As Gershenson et al. (2022) point out, if the effects of Black teachers on student outcomes are explained by a set of mindsets, practices, and skills they bring to their work, then it *may* be possible to train the mostly-White teacher workforce in these areas. Alternatively, if effects are driven by role modeling, then the only real option is to engage in different approaches to recruitment and retention of Black and other individuals of color to substantially alter the demographics of the population of public-school teachers.

Motivating Literature

Although the research literature linking teacher-level characteristics to student outcomes has crystallized around the benefit of same-race/ethnicity matching, less is known from this same research tradition about the mechanisms driving these effects. Theory, largely grounded in sociological and human development perspectives, suggests three possible pathways. First, Black and other historically marginalized and minoritized students of color benefit from having teachers who look like them as role models, particularly given the way in which their career and training exemplifies academic success (Villegas & Lucas, 2004). Seeing a more equitable distribution of power in schools—relative to society more broadly—can help draw students of color into the classroom environment and build a sense of self-efficacy and belonging (Bristol & Martin-Fernandez, 2019).

Second—and not mutually exclusive from the first pathway—Black and other teachers of color may be better equipped than White teachers at teaching students of color. White teachers are not inherently unable to teach students of color, but may be more likely to adopt and maintain deficit views and colorblind ideologies that presume that individual factors—rather than systemic racism—are responsible for the academic challenges that students of color may experience (Lewis, 2001). In contrast, Black teachers who situate high expectations for academic success at the "base" of instruction (Ladson-Billings (1995a, p. 160) can help offset "stereotype threat" and the risk of

confirming a negative stereotype about a group (Steele & Aronson, 1995). The culturally relevant pedagogy of Black teachers also is described as one of "opposition", where teachers use their understanding of students' culture and lives to guide instruction (i.e., "cultural competence"), as well as support students to critique cultural norms, values, and institutions that produce and maintain social inequities (i.e., "critical consciousness"; Gay, 2000; Ladson-Billings, 1995b; Paris, 2012).

Third, Black and other teachers of color may be better at teaching *all* students. While race-conscious instruction is central to Black scholars' definition and description of culturally relevant pedagogy, Ladson-Billings (1995a) also argues that the "pedagogical excellence" of Black teachers includes additional features of "good teaching": differentiating instruction to meet the needs of individual students as they pursue academic goals, ensuring that classrooms are well-organized for learning without creating an exclusionary climate, and building strong interpersonal relationships with students to support engagement in the classroom environment. Gay (2000) similarly describes "the power of caring" as a key component of culturally responsive teaching. Potentially driven by these practices, Asian, Black, Hispanic, and White students report feeling better cared for and more academically challenged when they have a teacher of color (Cherng & Halpin, 2016).

I summarize this conceptual framework in Figure 1, which links Black teachers to outcomes of Black and non-Black students through both role-modeling channels and specific mindsets and pedagogical practices. In the framework, Black teachers are hypothesized to improve components of students' social-emotional learning (SEL) first, potentially translating into increased school attendance and ultimately into test scores. Black students are most likely to benefit from Black teachers—inclusive of both role-modeling and pedagogical practices—though non-Black students can benefit too—primarily through Black teachers' mindsets and practices. The conceptual framework further lays the groundwork for the current analyses by identifying the measures that are or are not visible in the available data, as well as the pathways that can be tested causally versus in an exploratory way.

Despite rich theoretical discussion and qualitative exploration on why Black teachers matter, quantitative scholars generally have been quite limited by available data to explore these mechanisms and mediating pathways in any rigorous way. Exploiting the random assignment of teachers to students in the Project STAR/class size experiment in Tennessee from the 1980s, Gershenson et al. (2022) argue in favor of the role-modeling hypothesis given that Black teachers only impacted the test scores, absences, and long-run educational attainment of Black students and not White students. Further, these effects persisted even when accounting for observable background characteristics of teachers (i.e., experience, highest degree attained, status on a career ladder) that the authors suggest may be proxies for good teaching. Edmonds (2022) makes a similar argument in a non-experimental study by estimating effects on test scores and suspensions.

At the same time, making a case for role-modeling effects—which are difficult to observe directly—generally requires ruling out other possible explanations and, thus, demands fairly detailed data on both teachers and students. The background characteristics of teachers available in the large administrative datasets used by Edmonds (2022) and Gershenson et al. (2022) do not align with the theoretical literature on teaching pedagogy that emphasizes specific mindsets and practices. These studies also focus on test scores, school behaviors, and educational attainment measures, which could be influenced by role-modeling channels but likely only insomuch as students first develop positive school experiences and a sense of belonging in the classroom (Bristol & Martin-Fernandez, 2019). A growing number of studies examine links between teacher-student race/ethnicity-matching and SEL measures, including engagement, motivation, and social ties (e.g., Egalite & Kisida, 2018; Rasheed et al., 2020; Wright et al., 2017). However, none of these studies can support robust causal claims through experimental designs, and they do not examine teacher-level perceptions, expectations, and practices that may drive these effects.

To my knowledge, this study is one of just a handful of experiments to estimate the effects of

Black teachers on student outcomes (Constantine et al., 2009; Dee, 2004; Gershenson et al., 2022), and the only experiment to incorporate student SEL measures and teacher mindset and practices.

Data

Sample and Experimental Design

The sample and data for this study come from a research project called the National Center for Teacher Effectiveness (NCTE), which examined characteristics of effective teachers and effective teaching in upper-elementary classrooms (i.e., fourth and fifth grade). The four anonymized partner districts all are urban contexts, geographically located in the Northeast, Mid-Atlantic, and Southeast regions of the U.S. The project was interested primarily in math instruction and so some of the teacher and student measures focus on this content area. At the same time, all participating teachers were generalists who taught all core subjects, suggesting that measures may generalize across teachers' work and instruction. Other studies show that teachers who are effective in math—as measured by valued-added to test scores and classroom observations—also tend to be effective in English language arts (Cohen, Ruzek, & Sandilos, 2018; Goldhaber, Cowan, & Walch, 2013).

In academic year 2012-13, which is the final year of the three-year study, the project conducted an experiment that randomly assigned teachers to class rosters within schools. The research team worked with district and school leaders to identify schools and school-grade combinations that were eligible for random assignment, meaning that there were at least two teachers in each school-grade and principals considered the set of teachers as capable of teaching any of the rosters of students that they (or their leadership team) created. Out of 91 eligible teachers in the partner schools in the 2012-13 school year, 71 were randomly assigned to rosters (n = 1,283 students) and constitute the main sample for this analysis. [1]

[1] I exclude four intact randomization blocks with 10 teachers who originally agreed to participate in the random assignment study but were missing relevant data for one of several reasons: four teachers left the study before the beginning of the 2012-13 school year for reasons unrelated to the experiment (i.e., leaving the district or teaching, maternity leave, change

Students and teachers in the experiment look similar to the larger NCTE sample, broader populations in the partner districts, and urban school districts across the U.S. (see Appendix Table 1). Over 35% of students are Black and 25% Hispanic, with at least two-thirds of students eligible for free or reduced-price lunch (FRPL). Roughly one-fourth of teachers are Black and 70% White. Black versus White teachers in the experiment also have similar background characteristics to each other, including undergraduate major in education, certification pathway, advanced degree, and teaching experience ($p = 0.396$ on joint test of significance; see Supplemental Table 1).[2] The moderately sized experimental sample includes only a handful of Asian and Hispanic teachers, who I keep in the analysis to ensure fidelity of the randomized design. However, I cannot draw strong inferences about these groups.

Measures

Student Outcomes. I examine effects of Black versus White teachers on six student outcomes. Three SEL measures come from a student survey administered in the spring as part of the NCTE project (see Appendix Table 1 for descriptive statistics, Supplemental Table 2 for item text, and Blazar & Kraft, 2017 for exploratory factor analyses): (i) student-reported *Self-Efficacy* captures students' effort, initiative, and perception that they can complete tasks (10 items, internal consistency reliability [α] = 0.76); (ii) *Engagement and Happiness in Class* asks students about their affect, happiness in, and enjoyment of class activities (5 items, $\alpha = 0.82$); and (iii) *Self-Regulation* captures the extent to which students regulate their behavior to align with teachers' expectations (3 items, $\alpha = 0.74$).

in teaching assignment); the principal of two teachers decided that it was not possible to randomly assign rosters to these teachers; and four teachers had random assignment partner(s) who left the study for either of the two reasons above. As randomization blocks are analogous to individual experiments, dropping individual ones does not threaten the internal validity of results.

[2] While I cannot reject the null hypothesis that background characteristics jointly differ between Black and White teachers, it is notable that that all of the Black teachers in experiment are female, compared to 80% of White teachers. White teachers also outperformed Black teachers on a test of math knowledge that includes items from teacher licensure exams, potentially reflecting racial biases in these sorts of exams demonstrated in other studies (Goldhaber & Hansen, 2010). However, the analyses in this study show that Black teachers outperform White teachers on several other metrics, including their mindsets, instructional practices, and impacts on student outcomes.

District records include three additional outcomes: state assessments in (iv) math and (v) English language arts (ELA); and (vi) absences from school. While the surveys were available in just one year, district records were available during the experiment, prior school years, and all subsequent years through 2018-19 (i.e., the year before Covid interrupted district data collection). Therefore, I estimate effects on measures captured at the end of the experimental year when students were in elementary school and up to six years later in high school.[3] I standardize test scores and survey responses to have a mean of 0 and a SD of 1. Given the skewed nature of the absence data (see Appendix Table 1), I take the natural log of absences plus 1. Correlations between the student-level measures show that SEL dimensions, absences, and test scores all correlate with each other and across time points (see Supplemental Table 3).

Teacher Mindsets and Practices. The NCTE project also collected a range of teacher-level measures, both from a teacher survey administered in the fall and from videotaped lessons. Teachers contributed an average of three lessons per school year, which trained raters scored on the Classroom Assessment Scoring System (CLASS) observation instrument (Pianta et al., 2012). Across these two sources of data, I focus on five total measures that align to theoretical discussion on "good" teaching practices and mindsets likely to show up in the classrooms of Black teachers. The survey includes three measures (see Appendix Table 1 for descriptive statistics, and Supplemental Table 4 for survey item text): (i) teacher-reported *Growth Mindset Beliefs* captures the extent to which teachers view student intelligence as malleable versus fixed (7 items, $\alpha = 0.82$), which aligns with Ladson-Billings (1995b) noticing that culturally relevant teachers hold and then act on beliefs that "knowledge is not static" (p. 481); (ii) *Preparation for Instruction* (15 items, $\alpha = 0.78$) identifies the amount of time teachers spend

[3] For longer-run test scores and absences, I use measures captured in the most recent year/grade level available for each student, as long as they were enrolled in high school in that most recent year. District records also include student suspensions, which I include as a control measure because it explains residual variation in outcomes. However, I do not include suspensions as an outcome given that only 4% of students received any suspensions and most teachers had zero students who were suspended.

planning for instruction and collecting formative assessment data—some items specific to math and some not—as well as the extent to which they use this information to deliver differentiated instruction that attends to individual students' needs (for connections to literature on culturally responsive teaching, see Kieran & Anderson, 2019; Ladson-Billings, 1995a); and (iii) *Relationships with Students and Families* (4 items, $\alpha = 0.63$) includes the rapport teachers develop with students in and outside of the classroom, and the amount of time teachers spend talking with parents and families about students' learning and behavior (Gay, 2000).[4]

Observations of classrooms include two additional measures (see Supplemental Table 5 for item text, and Blazar et al. 2017 for exploratory and confirmatory factor analyses)[5]: (iv) *Classroom Support* focuses on teachers' interpersonal relationships with students around classroom activities and content, including creating a positive classroom climate, and teachers' sensitivity to and respect for student ideas and perspectives (9 items, $\alpha = 0.90$, adjusted intraclass correlation of between versus within teacher-variation [ICC] = 0.63); and (v) *Classroom Organization* captures teachers' behavior management skills and the extent to which teachers' approach to addressing student (mis)behaviors

[4] In order to align the teacher survey constructs to the theoretical literature on culturally responsive teaching, I make a couple of theory- and data-driven changes to the originally designed constructs. The *Preparation for Instruction* items were developed to capture two constructs: out-of-class preparation and formative assessment, each with its own response scale (see Supplemental Table 4). However, exploratory factor analyses suggest that items cluster together to form a single construct. Theory on culturally relevant and culturally responsive teaching also describes both types of practices as jointly facilitating teachers' knowledge of students and then delivery of student-oriented classroom instruction (Kieran & Anderson, 2019). Second, one item from the *Relationships with Students and Families* construct—focused on teachers' interactions with family members—uses a different survey stem and response scale than the other items. Unsurprisingly, reliability is higher when excluding this item ($\alpha = 0.72$). I include the item given guidance from the theoretical literature. Aligned to this discussion, the predictive power to student outcomes generally is stronger when including the additional item.

[5] Video capture occurred with a freestanding, three-camera, digital recording device and lasted roughly 45 to 60 minutes. One camera focused on the front of the classroom, while two others focused on student tables. Two microphones—one attached to the recording device and another worn by the teacher—picked up classroom talk. Teachers were allowed to choose dates, but were directed to select typical lessons and exclude days when students were taking a test. Although it is possible that these lessons captured instructional practice that were unique from a teachers' general instruction, teachers did not have any incentive to select lessons strategically. Following protocols outlined by CLASS developers, I calculated teacher-level scores for each measure by averaging scores across each 15-minute segment in a given lesson and across items within the dimension. Then, to minimize measurement error and account for variation in the number of lessons teachers contributed to the dataset—with an average of three per school year—I calculated Empirical Bayes estimates of teacher-level scores that are shrunken back to the mean based on their precision.

avoids creating a negative classroom culture (3 items, $\alpha = 0.72$, ICC = 0.47); this measure aligns with scholars' noticing that the behavior, physical movements, and language of minoritized students often is misunderstood (Fenning & Rose, 2007; Gay, 2000). Reliability indices are similar to other large-scale video studies of classroom instruction (e.g., Kane & Staiger, 2012).

I standardized all teacher-level measures to have a mean of 0 and a SD of 1. Correlations between the teacher mindset and practice measures follow expected patterns (see Supplemental Table 6). For example, outside observers' assessment of teachers' *Classroom Support* is most strongly correlated with teacher-reported *Preparation for Instruction* ($r = 0.37$) and *Relationships with Students and Families* ($r = 0.22$).

Empirical Strategy

Average or Total Effects of Black versus White Teachers on Student Outcomes

The randomized design allows for a straightforward approach to estimate the average or total effect of Black versus White teachers on student outcomes. I begin with the following model:

$$Y^{c}_{isgj(t+n)} = \beta_0 + \beta_1 BlackTch_{ij(t=0)} + \beta_2 AsianOrHispanicTch_{ij(t=0)} + \beta_3 X_{i(t=0)} + \beta_4 Z_{j(t=0)} + \gamma_{sg} + \varepsilon_{isgjt} \quad (1)$$

where $Y^{c}_{isgj(t+n)}$ is the set of short- and longer-run outcomes *c* for student *i* in school *s* and grade *g* randomly assigned to teacher *j*. The time subscripts $t + n$ indicate that outcomes are captured at the end of the year working with the teacher, as well as up to six years later when students were in high school (for test scores and absences). The main independent variable, $BlackTch_{ij(t=0)}$, captures whether or not students' randomly assigned teacher is Black. I condition on whether or not students' randomly assigned teacher is Asian or Hispanic (combined into one group, given the limited sample size for each), leaving White as the left-out category. To match the blocked randomized design, I control for school-grade fixed effects, γ_{sg}. In my preferred models, I also condition on a set of

background student characteristics, X_i, to increase precision (see Appendix Table 1 for full list). Controlling for background teacher characteristics, Z_j, also helps increase precision, as well as address the fact that randomization assigned these characteristics to students at the same time that it randomized teacher race/ethnicity. I calculate robust standard errors clustered at the teacher level to account for the clustered randomized design, with students nested within teachers' classrooms.

Estimates from equation (1) provide evidence of the effect of Black versus White teachers, on average across students. I probe the differential effect of Black teachers on the outcomes of Black students (i.e., race-matching effects) versus on non-Black students by dividing the sample, re-estimating effects for each subgroup, combing the variance-covariance matrices, and conducting post-hoc Wald tests of coefficient equivalence. These subgroup analyses provide insight into whether Black teachers are more effective than White teachers overall, or whether Black teachers serve as role models or engage in pedagogies that are uniquely beneficial to Black students.

The internal validity of resulting estimates rests on two assumptions, both of which are met in this study (see Appendix Table 2). First, I confirm baseline balance by showing that pre-treatment student characteristics are unrelated to the race of their randomly assigned teacher ($p = 0.313$ on joint test of significance). Second, I show that non-compliance (i.e., students moving out of their randomly assigned teachers' classroom) and missing data do not lead to imbalanced groups ($p = 0.466$ on joint test of significance for non-compliance, and $p = 0.103$ to 0.397 across tests for missing data on each data source).[6]

[6] While 31% of students moved out of their randomly assigned classroom, this is not a concern on its own given the intent-to-treat empirical framework. I estimate the effect of students' randomly assigned teacher, not the teacher with whom they eventually worked. Seventeen percent of non-compliers moved schools within the same district, 21% moved out of the district, and 62% switched classrooms within the same school (though in most cases to another teacher that was not part of the research study). The compliance rate in this study (69%) is higher than in other recent studies that randomly assigned teachers to classes (e.g., 66% to 27% across the six districts including in the Measures of Effective Teaching project; Kane et al., 2013), but still poses a threat to internal validity if there is differential attrition from the dataset. For example, 79% of non-compliers and 30% of the full experimental sample did not participate in primary data collection of student surveys. Importantly, analyses presented in Appendix Table 2 indicate that missing data is not a threat to internal validity because the characteristics of teachers' students remain balanced. Given that most non-compliers are

Mediating Pathways

In a set of exploratory analyses, I further examine two types of mediation outlined in the conceptual framework. The first examines whether the effect of Black versus White teachers on proximal outcomes—such as components of students' SEL—mediate effects on more distal student outcomes—such as test scores—using the following equation:

$$Y^{c}_{isgj(t+n)} = \omega_0 + \omega_1 BlackTch_{ij(t=0)} + \omega_2 AsianOrHispanicTch_{ij(t=0)} +$$

$$\omega_3 Y^{-c}_{isgj(t)} + \omega_4 X_{i(t=0)} + \omega_5 Z_{i(t=0)} + \gamma_{sg} + \varepsilon_{isgjt} \quad (2)$$

$Y^{-c}_{isgj(t)}$ is a vector of student outcomes that precede $Y^{c}_{isgj(t+n)}$ in the conceptual framework. For example, effects of Black teachers on students' SEL and on school attendance may drive later effects on test-score performance. By controlling for the proximal outcome, I can parse the direct effect of Black versus White teachers on more distal outcomes from the effect of the hypothesized mediator.

Following a mediation framework, the difference between the *direct effect* of Black versus White teachers from equation (2), ω_1, and the *average* or *total effect*, β_1, from equation (1) is interpreted as the *mediated* or *indirect effect* (VanderWeele, 2015). If this difference is zero, then there is no mediating effect. If the difference is large and negative, this is indication of mediation because the more proximal student outcomes reduce the magnitude of the total effect and, thus, explains some of the effect of Black versus White teachers on the more distal outcome. In the results presented below, I consider the percent of the total effect that is explained by the mediated effect, or ω_1 minus β_1 divided by β_1.

Similarly, I assess mediating pathways that run through the set of observed teacher mindset and practices by including these measures in place of $Y^{-c}_{isgj(t+n)}$. I further interact the teacher mindset

missing data, it is not possible to estimate treatment-on-the-treated using an instrumental variables technique. In the validity tests, coefficients on prior math and ELA achievement often are statistically significant. However, coefficients are roughly equal and opposite in magnitude because test scores are highly correlated across content areas (r = 0.63; see Supplemental Table 3). When only including one prior test score, coefficients are not statistically significantly different from zero in any model.

and practice measures with teacher race/ethnicity groups, as the relationship between a given mediator and outcomes may differ by teacher race/ethnicity. This setup often is referred to as moderated mediation (VanderWeele, 2015). Ten teachers in the analysis sample were missing survey and video data (n = 2 Black and 8 White teachers). However, teachers with and without mediator data do not differ in their background characteristics, either tested individually or as a group (p = 0.309 on joint test of significance). Rather than drop full classrooms and randomization blocks, I impute missing values with predicted values from a regression model of the teacher mindset and practice measures on background teacher characteristics and randomization block fixed effects.

The mediation analyses are only suggestive of the mechanisms that drive effects of Black versus White teachers on student outcomes because of potential biases caused by unobserved confounders and the failure of "sequential ignorability" (Imai et al., 2011). Because it is not possible to randomly assign self-efficacy, engagement, self-regulation, absences, etc. to students, these student-level mediators are themselves affected by other unobserved factors (e.g., parental involvement). Similarly, it is not possible to randomly assign teacher mindset and practice measures to teachers. At the same time, in this study, students were randomly assigned to teachers who vary not only in their race/ethnicity but also in the knowledge and skills they possessed up until that point. Therefore, to limit potential biases in the teacher-level mediation analyses, I focus on mindset and practice measures captured in years prior to the experiment.

Results

Average or Total Effects of Black Teachers on Student Outcomes

In Table 1, I present estimates of the average or total effect of Black versus White teachers on the set of short- and longer-term student outcomes, on average across students and in subgroups of

Black and non-Black students.[7] Like Gershenson et al. (2022) and Edmonds (2022), I find that teacher background characteristics do not change estimates meaningfully. The same is true for inclusion versus exclusion of student characteristics, which is a useful test of internal validity.

I find very large effects of Black versus White teachers on intrapersonal components of students' SEL that are localized to Black students. Assignment to a Black versus a White teacher increases Black students' self-reported *Self-Efficacy* by 0.85 SD and their *Engagement and Happiness in Class* by 0.69 SD. For both outcomes, effects of Black versus White teachers for Black students are statistically significantly larger than effects for non-Black students. The magnitude of the coefficient of the effect of Black teachers on the *Self-Regulation* of their Black students is potentially meaningful (0.22 SD), but not statistically significantly different from zero.

Expanding to other outcome measures, I find that Black teachers have large effects on short-term absenteeism of all students, though the effect is larger for Black students (47% decrease, exponentiating the coefficient that is estimated in log units) compared to non-Black students (22% decrease). Effects on Black students' absences persist to a large degree up to six years later when students are in high school (43% decrease). In the short term, Black teachers also improve the math test scores of both Black and non-Black students (0.24 SD, on average across students, with no differential effect between subgroups). None of the effects on math achievement in high school are statistically significantly different from zero. However, point estimates suggest persistence over time for Black students (0.12 SD) but not non-Black students. The reverse is true for effects on ELA achievement. In the short term, Black teachers have very large impacts on the ELA test scores of Black students (0.38 SD), with potential persistence over time (0.17 SD, though not statistically significantly different from zero). For non-Black students, short-term effects are positive but small,

[7] For transparency, I report effects for Asian or Hispanic versus White teachers—which are estimated in the model—in Supplemental Table 7. Patterns are very similar to the effects of Black versus White teachers. However, I do not focus on these findings given the very small sample of Asian and Hispanic teachers and students.

while effects in high school are larger (0.07 versus 0.22 SD, not statistically distinguishable).

The large number of estimates presented from the same randomization could result in seeing false positives due to multiple hypothesis testing. To address this concern, I consider a Benjamini-Hochberg (1995) adjustment that accounts for the number of tests conducted (n = 63) and an allowable false discovery rate, which I set at 20%. The resulting critical value for statistical significance of 0.09 is above the standard threshold of 0.05. This is possible given that a large share of the estimates reported in Table 1 are statistically significant.[8] I further examine the robustness of the main findings to potential lingering concerns regarding differential attrition and missing data by re-estimating results across 10 multiply imputed datasets. Magnitudes of estimates are very similar to the main findings, but often estimated less precisely (see Supplemental Table 8).

Mediating Pathways

Next, I explore the extent to which effects of Black versus White teachers on proximal student outcomes mediate effects on more distal outcomes (Table 2), as well as potential mediation through teacher mindset and practice measures aligned to "good teaching" (Table 4). In both tables, estimates are percent changes in the total effect (see Table 1) after accounting for a given mediator. (See Supplemental Tables 9 and 10 for direct effect estimates used to calculate percent change, for student- and teacher-level mediation analyses, respectively.) The sample sizes are slightly different between Tables 2 and 4 because, in student-level mediation analyses, I construct a consistent sample of students who have data on each outcome and all relevant mediators.

Following the conceptual framework and the main results, I allow mediating pathways to vary between Black and non-Black students. But, I exclude several mediating pathways when the average or total effects are not statistically significant and below 0.1 (in absolute value). The technical literature

[8] Following Benjamini-Hochberg (1995), I start by rank ordering all p-values from smallest to largest. Then, I use the formula $i/m*Q$, where i is the rank of the p-value, m is the total number of tests (n = 63), and Q is the false discovery rate that I set at 0.2. Finally, I identify the largest p-value that is less than the adjusted critical value.

on causal mediation argues that mediation only can occur when there is a statistically significant average or total effect (VanderWeele, 2015). I set a slightly lower bar (at the 0.1 effect size threshold), given the moderate sample size and the exploratory nature of the mediation analyses. Readers should interpret these results with caution. Percent change estimates can get quite large when the total effect is small and estimated with noise, as the total effect is the denominator in the percent change equation. Cells are highlighted in gray scale to identify larger versus smaller degrees of mediation.

Results in Table 2 suggest that the effects of Black versus White teachers on the *Self-Efficacy* and *Engagement* of their Black students meaningfully mediate effects on all subsequent outcomes. Each measure mediates the effect of the other, with large percent change estimates (54% and 81% of the total effect explained by the mediator) because *Self-Efficacy* and *Engagement* are strongly correlated (r = 0.65; see Supplemental Table 3) and because Black teachers have large effects on both outcomes. These two measures also mediate effects on Black students' *Self-Regulation*, with percent change estimates above 100%. This is one instance where a smaller total effect that is substantively meaningful but not statistically significant may lead to large percent change estimates. Effects of Black teachers on *Self-Efficacy* and *Engagement* further mediate effects on short- and longer-term absences (13% and 10%), as well as short-term math test scores (over 100%) and longer-run math and ELA test scores (43% to 58%). Short-term effects on absences mediates the largest shares of longer-run effects on absences (39%), as well as the largest share of longer-run effects on math test score (59%). For non-Black students, I observe very little to no mediation through their *Self-Efficacy* and *Engagement*, which is largely mechanical: Black teachers do not impact these outcomes of non-Black students. In contrast, short-term effects of Black teachers on the test scores of non-Black students mediate a very large share of effects on longer-run test scores (over 70%). This finding may explain why the effect of Black teachers on non-Black students' ELA achievement grows over time.

To explore teacher-level mediation, I first show differences in the mindset and practice

measures between Black and White teachers (Table 3). Given the exploratory nature of these analyses, I present estimates for the experimental sample and the broader sample of teachers from the NCTE project to inform generalizability. I estimate these between-group differences using the same set of controls used in the student-level regressions, as the teacher mindset and practice measures also are included in the student-level analyses. Black teachers outperform their White colleagues on most measures, though differences are not always statistically significant given the limited sample size. In the experimental sample, between-group differences are largest for *Classroom Support* (0.51 SD, preferencing the model with school-grade fixed effects and background student and teacher characteristics) and *Classroom Organization* (0.58 SD). Estimates for *Growth Mindset Beliefs* and *Preparation for Instruction* also are substantively meaningful (roughly 0.2 SD in the preferred model) but not statistically significant. The pattern is reversed in the non-experimental sample, with large between-group differences on *Growth Mindset Beliefs* and *Preparation for Instruction* (0.5 to 0.8 SD) and smaller but still meaningful differences for *Classroom Support* and *Classroom Organization* (0.2 to 0.4 SD). In both samples, differences between Black and White teachers on *Relationships with Students* are smaller and more sensitive to control set.

In Table 4, I find that teacher reports of *Preparation for Instruction* is a fairly consistent mediator of the effect of Black versus White teachers across a range of student outcomes, for Black and non-Black students, but primarily in the short term. This pattern makes sense, given that preparing for and differentiating instruction can immediately change classroom experiences in the short term. *Preparation for Instruction* mediates a sizeable portion of the effect of Black teachers on Black students' *Self-Efficacy* (42%), *Engagement* (26%), short-run absences (11%) and math test scores (24%). I observe similar patterns of mediation for non-Black students. In contrast, *Growth Mindset Beliefs* is a strong mediator of longer-run outcomes (i.e., 73% for Black students' ELA achievement), with more moderate mediation in the short term (e.g., 39% for this same outcome). It may be that teacher mindsets take

longer to influence students' own mindsets and beliefs. *Classroom Organization* is a strong mediator of both short- and longer-run outcomes (e.g., roughly 40 to 50% for Black students' short-term SEL measures, and over 100% for Black students' short- and longer-run math achievement). That said, percent change estimates go in the "wrong" direction (i.e., positive) for short-term absences, potentially reflecting a tradeoff between organizational and engagement-oriented classroom practices. Indeed, the *Classroom Organization* of Black teachers is a very strong predictor of Black students' *Self-Efficacy* (0.9 SD) but is associated with more short-term absences (see Supplemental Table 11 for estimates linking the teacher mindset and practice measures to student outcomes).

Finally, the two measures related to teacher-student relationships and rapport explain only a small share of the effect of Black versus White teachers on student outcomes and often just for non-Black students. One explanation is that differences between Black and White teachers on *Relationships with Student and Families* are small (see Table 3). Another explanation is that, even though Black teachers outperform White teachers on *Classroom Support*, this measure is associated with worse student outcomes in several instances (see Supplemental Table 11). This pattern similarly reflects a possible tradeoff between engagement-oriented versus other types of practices.

I describe the mediation analyses as exploratory because I cannot definitively rule out potential confounders. At the same time, both student- and teacher-level mediation analyses pass a robustness test proposed by Imai et al (2011) that examines whether unobserved factors that show up in the error term are correlated across the system of equations. More specifically, I correlate residuals from a model that regresses a given mediator on treatment and a model that regresses student outcomes on treatment and the same mediator. Imai et al (2011) argue that "sequential ignorability" holds when the correlation is zero. I find that correlations often are zero (to three decimal places) and no higher than 0.13 (in absolute value; see Supplemental Table 12). In a similar application of Imai et al's (2011) procedure, Tran and Gershenson (2021) interpret a correlation of -0.9 as "relatively small" because

correlations are bounded between -1 and 1 (p. 194).

Discussion and Conclusion

The results of this study generally confirm longstanding theory and qualitative inquiry regarding the importance of Black teachers to students' classroom experiences and outcomes, which—to my knowledge—have not been fully tested in quantitative and experimental research. First, I find that Black teachers have very large effects on components of their Black students' SEL (upwards of 0.85 SD). These effects compare quite favorably to other analyses of SEL-oriented interventions in schools (roughly 0.23 SD; for a meta-analysis, see Durlak et al., 2011), as well as to natural gains in these sorts of measures from one year to the next (roughly 0.02 to 0.03 SD for upper-elementary students; Soland et al., 2022). Mediation analyses further suggest that the effects on Black students' SEL likely translate into subsequent effects on short- and longer-term school attendance and test scores, with a remarkable degree of persistence over time particularly in the effects on absences. This pattern of persistence is almost unheard of in education research (Bailey et al., 2017).

Second, I find that Black teachers benefit all students—with meaningful decreases in absenteeism and increases in test scores—not just students who look like them. This finding is consistent with theory (Villegas & Lucas, 2004) and some exploratory quantitative research on student perceptions of teachers of color (Cherng & Halpin, 2016), but differs from prior experimental analyses showing that Black teachers impact Black students only (Gershenson et al., 2022). One explanation may be that most of the non-Black students in this study are Hispanic and Asian—reflecting demographic diversity across the four districts that come from the Northeast, Mid-Atlantic, and Southeast regions—while most of the non-Black students in the Project STAR experiment from Tennessee in the 1980s and analyzed by Gershenson et al. (2022) are White. Black teachers may benefit other minoritized students of color more than they impact White students. That said, when I limit the analysis sample to roughly 20% of the student sample that is White, I find some evidence of positive

effects of Black teachers on White students (e.g., 30% decrease in short-term absences), though precision is limited substantially (hence not shown or discussed in the main analyses). It also is possible that results may differ in other geographic contexts, with different demographic makeups and with different racialized histories of schooling.

Third, helping to explain the two patterns described above, I find suggestive evidence that the effects of Black versus White teachers likely are driven both by role-modeling channels *and* by the unique set of mindsets and practices that Black teachers bring to their work. I cannot directly observe and prove the existence of role modeling. However, the fact that Black teachers impact the self-efficacy and classroom engagement of their Black students but not for non-Black students points in this direction. As a psychological construct related to self-image and sense of self, role modeling may be most likely to impact these sorts of intrapersonal competencies.

I also provide evidence on the "pedagogical excellence" of Black teachers (Ladson-Billings, 1995b) that helps generalize beyond small-scale qualitative studies. The teacher-level mediating pathways are not always consistent, with most mediators associated with increases in some student outcomes but decreases in others. It also is puzzling that the two measures capturing teachers' relationships with students explain very little of the effects of Black teachers on student outcomes, given discussion of a "culture of caring" being so central to some definitions of culturally responsive teaching (Gay, 2000). Consistent with the theory, Black teachers *do* exhibit caring for their students, with scores on *Classroom Support* higher than those for White teachers. The fact that these scores are negatively associated with some student outcomes may seem counterintuitive, but aligns with other mixed-methods explorations in the same dataset (Blazar & Pollard, 2023). Instead, the teacher-level mediation findings focus attention on *Growth Mindset Beliefs*, *Preparation for Instruction*, and *Classroom Organization*, which all mediate sizeable shares of the effect of Black teachers on student outcomes. These patterns underscore Ladson-Billings' (1995a) argument for situating high expectations for

academic success at the "base" of instruction (p. 160).

Of course, there also is room to explore additional mediators, particularly for student outcomes where I observe only a moderate degree of teacher-level mediation (e.g., short-term absences). One place to focus is to measure and include "opposition" pedagogies that are central components of culturally responsive teaching and are thought to drive Black students' classroom engagement, but are not observed in this study. There may also be indirect paths that run through students' families and communities that also are not observed in this study in a comprehensive way. The extent to which Black teachers engage families in schooling activities may play a key role in explaining effects on student absences in upper-elementary grades, for example, where students likely have less control compared to their parents and guardians. Markowitz et al. (2020) show that, in the context of Head Start, teacher-child race/ethnicity matching is associated with increased parental engagement, as well as attendance.

Ultimately, the experimental and mediation analyses provide guidance for policy and practice related to teacher training and recruitment. The findings suggest that both approaches are necessary. At the same time, knowing what the policy and practice goals are does not make them simple or straightforward to achieve. On the recruitment side, increasing teacher diversity is a numbers problem at a bare minimum given very large demographic mismatches between Black teachers and Black students. In the full NCTE study, 22% of Black students had a Black teacher, compared to 83% of White students who had a White teacher. More challenging, the task of shifting demographics requires school systems to wrestle with systemic racism in workforce policies, including the fact that Black teachers were systematically ushered out of schools following school integration efforts in the mid 20th Century (Thompson, 2022) and current Black teachers often are underappreciated for their work (Griffin & Tackie, 2017). Further, while I concur with other scholars who advocate for training of the current teacher workforce that is mostly White (Gershenson et al., 2022), the skills that Black teachers

have more than White teachers, on average, and that benefit student outcomes may not be teachable, or at least not easily taught. Racial biases—instantiated, for example, in exclusionary discipline (e.g., Fenning & Rose, 2007)—are not easy to overcome.

Because of these challenges, I conclude by reiterating the fundamental points of this paper: Experimental evidence shows that assignment to a Black teacher produces some of the largest effects on student outcomes across all of the education research literature (Fryer, 2017). These effects show up in SEL, school attendance, and test scores. They persist over time, and they extend to Black and non-Black students alike. We must use these findings to compel large changes in policy and in practice.

References

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, *10*(1), 7-39.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*(1), 289-300.

Blazar, D., Braslow, D., Charalambous, C.Y. & Hill, H.C., (2017). Attending to general and mathematics-specific dimensions of teaching: Exploring factors across two observation instruments. *Educational Assessment*, *22*(2), 71-94.

Blazar, D., & Kraft, M. A. (2017). Teacher and teaching effects on students' attitudes and behaviors. *Educational Evaluation and Policy Analysis*, *39*(1), 146-170.

Blazar, D., & Pollard, C. (2023). Challenges and tradeoffs of "good" teaching: The pursuit of multiple educational outcomes. *Journal of Teacher Education*, *74*(3), 229-244.

Bristol, T. J., & Martin-Fernandez, J. (2019). The added value of Latinx and Black teachers for Latinx and Black students: Implications for policy. *Policy Insights from the Behavioral and Brain Sciences*, *6*(2), 147-153.

Cherng, H. Y. S., & Halpin, P. F. (2016). The importance of minority teachers: Student perceptions of minority versus White teachers. *Educational Researcher*, *45*(7), 407-420.

Cohen, J., Ruzek, E., & Sandilos, L. (2018). Does teaching quality cross subjects? Exploring consistency in elementary teacher practice across subjects. *AERA Open*, *4*(3).

Constantine, J., Player, D., Silva, T., Hallgren, K., Grider, M., & Deke, J. (2009). *An evaluation of teachers trained through different routes to certification.* NCEE 2009-4043. National Center for Education Evaluation and Regional Assistance.

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

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*(1), 405-432.

Edmonds, L. (2022). *Role models revisited: HBCUs, same-race teacher effects, and Black student achievement.* Working Paper.

Egalite, A. J., & Kisida, B. (2018). The effects of teacher match on students' academic perceptions and attitudes. *Educational Evaluation and Policy Analysis*, *40*(1), 59-81.

Fenning, P., & Rose, J. (2007). Overrepresentation of African American students in exclusionary discipline the role of school policy. *Urban Education, 42*(6), 536-559.

Fryer Jr, R. G. (2017). The production of human capital in developed countries: Evidence from 196 randomized field experiments. In *Handbook of economic field experiments* (Vol. 2, pp. 95-322). North-Holland.

Fryer Jr, R. G., & Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. *Review of Economics and Statistics*, *86*(2), 447-464.

Gay, G. (2000). *Culturally responsive teaching: Theory, research, and practice*. Teachers College Press.

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

Goldhaber, D., & 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., Cowan, J., & Walch, J. (2013). Is a good elementary teacher always good? Assessing teacher performance estimates across subjects. *Economics of Education Review*, *36*, 216-228.

Griffin, A., & Tackie, H. (2017). Through our eyes: Perspectives from black teachers. *Phi Delta Kappan*, *98*(5), 36-40.

Hernandez, D. J. (2011). *Double jeopardy: How third-grade reading skills and poverty influence high school graduation.* Annie E. Casey Foundation. Retrieved from: https://eric.ed.gov/?id=ED518818

Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review*, *105*(4), 765-789.

Jackson, C. K. (2009). Student demographics, teacher sorting, and teacher quality: Evidence from the end of school desegregation. *Journal of Labor Economics*, *27*(2), 213-256.

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

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

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains.* MET Project. Bill & Melinda Gates Foundation.

Kieran, L., & Anderson, C. (2019). Connecting universal design for learning with culturally responsive teaching. *Education and Urban Society, 51*(9), 1202-1216.

Ladson-Billings, G. (1994). Who will teach our children: Preparing teachers to successfully teach African American students. *Teaching Diverse Populations: Formulating a Knowledge Base*, 129-142.

Ladson-Billings, G. (1995a). But that's just good teaching! The case for culturally relevant pedagogy. *Theory into practice*, *34*(3), 159-165.

Ladson-Billings, G. (1995b). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, *32*(3), 465-491.

Lewis, A. E. (2001). There is no "race" in the schoolyard: Color-blind ideology in an (almost) all-white school. *American Educational Research Journal*, *38*(4), 781-811.

Markowitz, A. J., Bassok, D., & Grissom, J. A. (2020). Teacher-child racial/ethnic match and parental engagement with Head Start. *American Educational Research Journal*, *57*(5), 2132-2174.

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

Paris, D. (2012). Culturally sustaining pedagogy: A needed change in stance, terminology, and practice. *Educational Researcher*, *41*(3), 93-97.

Pianta, R. C., Hamre, B. K., & Mintz, S. (2012). *Classroom assessment scoring system: Upper elementary manual.* Teachstone.

Rasheed, D. S., Brown, J. L., Doyle, S. L., & Jennings, P. A. (2020). The effect of teacher–child race/ethnicity matching and classroom diversity on children's socioemotional and academic skills. *Child Development*, *91*(3), e597-e618.

Redding, C. (2019). A teacher like me: A review of the effect of student–teacher racial/ethnic matching on teacher perceptions of students and student academic and behavioral outcomes. *Review of Educational Research*, *89*(4), 499-535.

Rivkin, S. G. (1995). Black/white differences in schooling and employment. *Journal of Human Resources*, 826-852.

Soland, J., Rimm-Kaufman, S. E., Kuhfeld, M., & Ventura-Abbas, N. (2022). Empirical benchmarks for changes in social and emotional skills over time. *Child Development*, *93*(4), 1129-1144.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797.

Thompson, O. (2022). School desegregation and Black teacher employment. *Review of Economics and Statistics*, *104*(5), 962-980.

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

U.S. Department of Education. (2019). *The condition of Education.*

VanderWeele, T. (2015). *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press.

Villegas, A. M., & Lucas, T. F. (2004). Diversifying the teacher workforce: A retrospective and prospective analysis. *Yearbook of the National Society for the Study of Education*, *103*(1), 70-104.

Ware, F. (2006). Warm demander pedagogy: Culturally responsive teaching that supports a culture of achievement for African American students. *Urban Education, 41*(4), 427-456.

Wright, A., Gottfried, M. A., & Le, V. N. (2017). A kindergarten teacher like me: The role of student-teacher race in social-emotional development. *American Educational Research Journal*, *54*(1_suppl), 78S-101S.

Figures and Tables

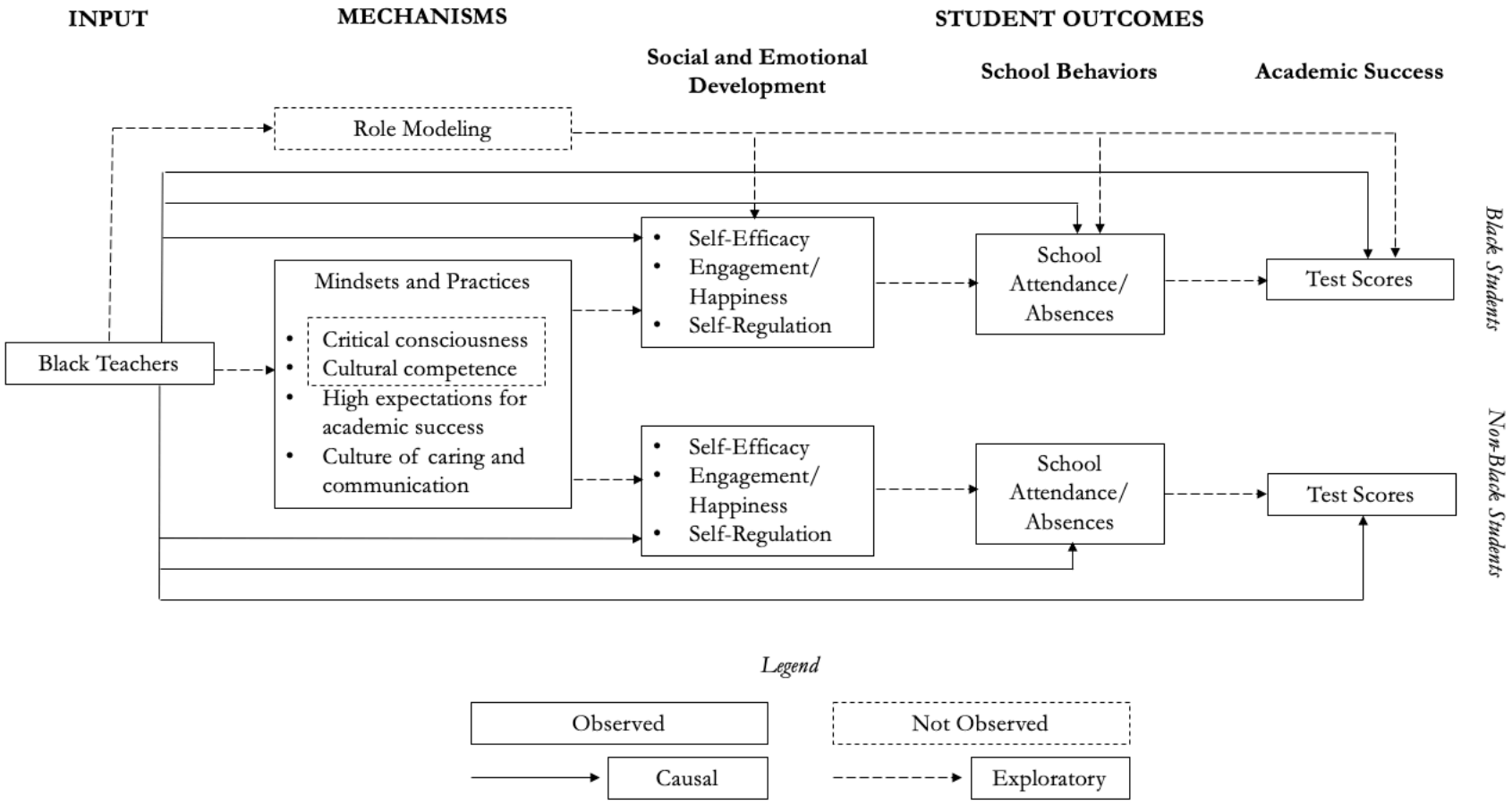

Figure 1. Conceptual framework linking Black teachers to student outcomes.

Table 1. Average Effects of Black versus White Teachers on Student Outcomes

Student Outcome	Sample	Average Effects Across All Students				Subgroup Effects		*p*-value on Diff.
						Black Students	Non-Black	
		(1)	(2)	(3)	(4)	(5)	(6)	
Panel A: End-of-Year Effects in Upper-Elementary School								
Self-Efficacy	903	0.337***	0.323***	0.385***	0.383***	0.853***	0.097	0.000
		(0.122)	(0.120)	(0.123)	(0.124)	(0.202)	(0.116)	
Engagement/Happiness	903	0.270**	0.236**	0.239	0.211	0.685***	-0.114	0.000
		(0.129)	(0.118)	(0.145)	(0.144)	(0.161)	(0.150)	
Self-Regulation	901	-0.040	-0.056	-0.035	-0.022	0.216	-0.024	0.199
		(0.121)	(0.117)	(0.110)	(0.104)	(0.157)	(0.109)	
Absences (log + 1)	1,017	-0.146**	-0.204***	-0.199***	-0.235***	-0.387***	-0.196***	0.051
		(0.065)	(0.049)	(0.070)	(0.049)	(0.100)	(0.041)	
Math Achievement	1,003	0.199**	0.161***	0.248***	0.237***	0.139	0.202***	0.639
		(0.082)	(0.055)	(0.088)	(0.065)	(0.126)	(0.071)	
ELA Achievement	1,004	0.091	0.114	0.127	0.193***	0.387***	0.074	0.032
		(0.112)	(0.068)	(0.103)	(0.061)	(0.133)	(0.058)	
Panel B: Follow-Up Effects in High School								
Absences (log + 1)	743	-0.041	-0.119**	-0.097	-0.137**	-0.355***	-0.054	0.063
		(0.090)	(0.057)	(0.083)	(0.057)	(0.118)	(0.091)	
Math Achievement	663	0.048	0.019	0.059	0.019	0.115	-0.003	0.410
		(0.101)	(0.079)	(0.110)	(0.073)	(0.106)	(0.096)	
ELA Achievement	670	0.135	0.136*	0.128	0.162*	0.174	0.216**	0.778
		(0.097)	(0.070)	(0.119)	(0.086)	(0.115)	(0.108)	
School-Grade Fixed Effects		X	X	X	X	X	X	
Background Student Characteristics			X		X	X	X	
Background Teacher Characteristics				X	X	X	X	

Notes: Estimates in each cell come from a separate regression model that regresses the outcome listed in each row on dummy indicators or whether or not students' randomly assigned teacher was Black, Asian or Hispanic (estimates not shown above; see supplemental materials), with White as the left-out category; and school-grade fixed effects. Some models include background teacher characteristics: gender, BA in education, traditional certification, MA degree, years of teaching experience, and math knowledge. Some models include background student characteristics include: gender, race/ethnicity, eligibility for free or reduced-price lunch, eligibility to receive special education services, limited English proficiency status, and prior-year absences, suspensions, and achievement in math and ELA. All outcomes other than absences are standardized. Robust standard errors clustered at the teacher level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2. Percent Change in Effects of Black versus White Teachers on Student Outcomes After Accounting for Student-Level Mediator

Student Outcome	Student Group	Sample	Self-Effiacy	Engage./ Happiness	Self-Regulation	Absences	Math Achieve.	ELA Achieve.
Panel A: End-of-Year Effects in Upper-Elementary School								
Self-Efficacy	Black	356	--	-54.2	-7.4	--	--	--
Engagement/Happiness	Black	356	-81.3	--	-10.0	--	--	--
Self-Regulation	Black	356	-112.0	-101.4	--	--	--	--
Absences (log + 1)	Black	356	-13.1	-9.5	-0.7	--	--	--
Absences (log + 1)	Non-Black	545	0.0	0.6	0.6	--	--	--
Math Achievement	Black	355	-102.5	-106.2	-18.5	-71.6	--	--
Math Achievement	Non-Black	545	-2.9	1.4	0.4	-5.0	--	--
ELA Achievement	Black	356	-7.6	-3.8	-6.1	-8.2	--	--
Panel B: Follow-Up Effects in High School								
Absences (log + 1)	Black	245	-17.5	-16.0	-4.3	-38.8	-5.6	-6.3
Math Achievement	Black	223	-57.7	-43.5	11.2	-58.8	-35.3	-30.6
ELA Achievement	Black	226	-42.9	-47.9	-2.5	-22.1	-25.8	-49.7
ELA Achievement	Non-Black	370	0.0	-17.5	-6.4	-14.3	-71.4	-69.8

Notes: "--" indicates the student outcome listed in each column is not used as a medaitor when predicting the outcomes in each row. The percent change estimates are calculated by dividing the direct effect of Black versus White teachers on student outcomes in models that include the mediator minus the total effect, divided by the total effect. Cells are highlighted in gray to reflect gradations of mediation: white for zero or positive percent change estimates, light gray for negative percent change greater than 0% and less than 10%, medium gray for negative percent change greater than or equal to 10% and less than 50%, and dark gray for negative percent change greater than 50%. Mediating pathways are excluded when the total effect is not statistically significant and less than 0.1 (in absolute value).

Table 3. Differences in Mindsets and Practices between Black and White Teachers

Teacher Mindset or Practice	(1)	(2)	(3)	(4)	(5)
	Panel A: Experimental Sample (N = 61)				
Growth Mindset Beliefs	0.452	0.386	0.390	0.212	0.223
	(0.308)	(0.237)	(0.236)	(0.266)	(0.265)
Preparation for Instruction	0.398*	0.095	0.092	0.189	0.190
	(0.209)	(0.117)	(0.117)	(0.145)	(0.145)
Relationships with Students and Families	0.476	-0.087	-0.087	-0.167	-0.164
	(0.382)	(0.153)	(0.150)	(0.203)	(0.199)
Classroom Support	0.656**	0.323*	0.324*	0.505**	0.506**
	(0.267)	(0.191)	(0.191)	(0.218)	(0.219)
Classroom Organization	0.505***	0.542***	0.545***	0.571***	0.575***
	(0.185)	(0.096)	(0.095)	(0.153)	(0.154)
	Panel B: Full Project Sample (N = 284)				
Growth Mindset Beliefs	0.393**	0.507**	0.471**	0.569**	0.508**
	(0.160)	(0.222)	(0.235)	(0.231)	(0.243)
Preparation for Instruction	0.395***	0.747***	0.784***	0.743***	0.799***
	(0.132)	(0.228)	(0.231)	(0.248)	(0.248)
Relationships with Students and Families	0.007	0.136	0.223	0.086	0.176
	(0.147)	(0.192)	(0.195)	(0.195)	(0.198)
Classroom Support	0.003	0.258	0.304	0.376	0.423
	(0.162)	(0.289)	(0.296)	(0.300)	(0.306)
Classroom Organization	-0.033	0.197	0.141	0.241	0.191
	(0.183)	(0.320)	(0.271)	(0.344)	(0.287)
School-Grade Fixed Effects		X	X	X	X
Background Student Characteristics			X		X
Background Teacher Characteristics				X	X

Notes: Estimates in each cell come from a separate regression model that regresses the teacher mindset or practice listed in each row on dummy indicators or whether or not the teacher is Black, Asian or Hispanic (not shown), with White as the left-out category. Some models include background teacher characteristics: gender, BA in education, traditional certification, MA degree, years of teaching experience, and math knowledge. Some models include background student characteristics include: gender, race/ethnicity, eligibility for free or reduced-price lunch, eligibility to receive special education services, limited English proficiency status, and prior-year absences, suspensions, and achievement in math and ELA. All outcomes other than absences are standardized. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4. Percent Change in Effects of Black versus White Teachers on Student Outcomes After Accounting for Teacher-Level Mediators

Student Outcome	Student Group	Sample	Growth Mindset	Prep. for Inst.	Relat. Stu./Fam.	Class Support	Class Org.
Panel A: End-of-Year Effects in Upper-Elementary School							
Self-Efficacy	Black	357	15.4	-41.9	-5.4	35.1	-47.1
Engagement/Happiness	Black	357	11.7	-25.7	-0.4	12.1	-38.5
Self-Regulation	Black	356	-6.5	-93.1	16.2	134.3	-44.4
Absences (log + 1)	Black	407	4.1	-10.6	0.8	-7.0	51.2
Absences (log + 1)	Non-Black	610	-26.0	4.1	-23.5	-39.8	43.9
Math Achievement	Black	401	36.0	-23.7	6.5	84.2	-122.3
Math Achievement	Non-Black	602	-33.7	-36.1	30.2	28.2	-58.4
ELA Achievement	Black	402	-39.0	5.9	-0.5	8.5	-11.9
Panel B: Follow-Up Effects in High School							
Absences (log + 1)	Black	283	63.9	34.7	19.4	50.1	-22.8
Math Achievement	Black	260	-12.2	41.7	51.3	98.3	-204.4
ELA Achievement	Black	262	-73.0	50.0	11.5	-46.0	-22.4
ELA Achievement	Non-Black	408	92.6	-22.7	7.9	31.5	13.0

Notes: The percent change estimates are calculated by dividing the direct effect of Black versus White teachers on student outcomes in models that include the mediator (not shown) minus the total effect, divided by the total effect. Cells are highlighted in gray to reflect gradations of mediation: white for zero or positive percent change estimates, light gray for negative percent change greater than 0% and less than 10%, medium gray for negative percent change greater than or equal to 10% and less than 50%, and dark gray for negative percent change greater than 50%. Mediating pathways are excluded when the total effect is not statistically significant and less than 0.1 (in absolute value).

Appendix

Appendix Table 1. Descriptive Statistics on Students and Teachers

	Experimental Sample		Full Project Sample			Full District Populations		
	Mean	SD	Mean	SD	*p*-value	Mean	SD	*p*-value
				Panel A: Students				
Demographics								
Female	0.50	0.50	0.50	0.50	0.975	0.50	0.50	0.992
Asian	0.10	0.30	0.07	0.26	0.005	0.09	0.28	0.084
Black	0.41	0.49	0.43	0.49	0.138	0.36	0.48	0.001
Hispanic	0.25	0.43	0.23	0.42	0.152	0.27	0.44	0.136
White	0.20	0.40	0.23	0.42	0.016	0.25	0.43	0.000
Other Race	0.05	0.21	0.04	0.20	0.268	0.04	0.20	0.254
Free or Reduced-Price Lunch	0.70	0.46	0.66	0.47	0.003	0.63	0.48	0.000
Special Education	0.07	0.26	0.12	0.33	0.000	0.14	0.35	0.000
English Language Learners	0.19	0.39	0.20	0.40	0.228	0.19	0.39	0.515
Outcomes: Upper-Elementary School								
Self-Efficacy (1 to 5)	4.09	0.93	4.08	0.94	0.845	--	--	--
Engagement/Happiness in Class (1 to 5)	3.99	0.90	4.09	0.86	0.001	--	--	--
Self-Regulation (1 to 5)	4.09	0.64	4.17	0.59	0.000	--	--	--
Absences (days)	6.08	7.08	6.36	7.05	0.187	6.01	6.47	0.749
Math Achievement (standardized)	0.027	0.944	0.044	0.93	0.542	0.01	0.99	0.514
ELA Achievement (standardized)	0.031	0.914	0.028	0.94	0.916	0.01	0.99	0.363
Outcomes: High School								
Absences (days)	10.30	15.49	12.77	18.87	0.000	--	--	--
Math Achievement (standardized)	0.049	0.931	0.057	0.915	0.831	--	--	--
ELA Achievement (standardized)	0.191	0.845	0.143	0.902	0.154	--	--	--
Students	1,283		12,532			175,472		
				Panel B: Teachers				
Demographics								
Female	0.79	0.41	0.81	0.39	0.733	0.76^	--	--
Asian	0.04	0.20	0.03	0.17	0.598	0.04^	--	--
Black	0.23	0.42	0.22	0.41	0.919	0.25^	--	--
Hispanic	0.03	0.17	0.03	0.17	0.982	0.07^	--	--
White	0.70	0.46	0.65	0.48	0.342	0.64^	--	--
Other Race/Missing	0.00	0.00	0.08	0.27	0.000	0.01^	--	--
Teacher Training, Experience, and Effectiveness								
BA in Education	0.55	0.50	0.53	0.50	0.818	--	--	--
Traditionally Certified	0.92	0.26	0.85	0.36	0.057	--	--	--
MA Degree	0.78	0.42	0.76	0.43	0.772	--	--	--
Teaching Experience (years)	11.07	6.39	10.21	7.12	0.320	--	--	--
Math Knowledge (standardized)	0.069	0.942	0.004	0.924	0.603	--	--	--
Math Value-Added (pre-experiment; student-level SD)	0.003	0.129	0.015	0.137	0.485	0.000	0.132	0.877
Mindsets and Practices								
Growth Mindset Beliefs (1 to 6)	4.45	0.95	4.38	0.93	0.626	--	--	--
Preparation for Instruction (1 to 5)	4.11	0.55	3.86	0.60	0.001	--	--	--
Relationships with Students and Families (1 to 5)	3.29	0.66	3.27	0.57	0.823	--	--	--
Classroom Support (1 to 7)	6.53	0.24	6.41	0.39	0.002	--	--	--
Classroom Organization (1 to 7)	4.09	0.33	4.12	0.42	0.571	--	--	--
Teachers	71		310			3,551		

Notes: -- indicates that data is not available from the full district population. ^ indicates that teacher demographic data for full district populations are pulled from publicly available information rather than micro data on teachers; these statistics capture demographics of teachers across all grade levels, rather than the upper-elementary sample for whom math value-added is calculated. P-values compare the characteristics of the the experimental sample to the full NCTE sample or the full district populations.

Appendix Table 2. Internal Validity Assumptions

Baseline Student Characteristics (interacted with non-compliance or missing data indicators in columns 2 through 7)	Baseline Balance	Non-Compliance	Missing Data				
			Survey (end-of-year)	Absences (end-of-year)	Test Scores (end-of-year)	Absences (high school)	Test Scores (high school)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	0.003	0.008	-0.020	0.007	0.016	-0.005	0.003
	(0.012)	(0.024)	(0.031)	(0.036)	(0.035)	(0.025)	(0.023)
Asian	0.036	0.055	0.057	0.110	0.111	0.117**	0.099**
	(0.027)	(0.045)	(0.058)	(0.074)	(0.069)	(0.045)	(0.040)
Black	-0.032**	0.007	-0.050	-0.065	-0.065	0.007	0.008
	(0.014)	(0.035)	(0.050)	(0.061)	(0.056)	(0.033)	(0.032)
Hispanic	-0.026	0.043	-0.011	-0.003	0.002	-0.001	-0.027
	(0.019)	(0.051)	(0.056)	(0.070)	(0.065)	(0.039)	(0.039)
White	-0.059	0.068	-0.017	0.024	-0.009	-0.006	0.059
	(0.038)	(0.052)	(0.079)	(0.100)	(0.104)	(0.063)	(0.053)
Free or Reduced-Price Lunch	-0.001	-0.011	-0.041	-0.030	-0.030	0.004	-0.005
	(0.010)	(0.041)	(0.042)	(0.060)	(0.059)	(0.043)	(0.042)
Special Education	-0.025	-0.030	0.159**	0.149	0.128	0.067	0.099
	(0.041)	(0.090)	(0.073)	(0.108)	(0.094)	(0.068)	(0.063)
Limited English Proficiency	0.013	0.033	0.095	0.042	0.030	-0.008	0.043
	(0.025)	(0.062)	(0.058)	(0.074)	(0.066)	(0.032)	(0.031)
Prior Absences (log + 1)	0.012	0.050*	0.013	0.060	0.084	-0.016	0.001
	(0.008)	(0.029)	(0.032)	(0.081)	(0.056)	(0.012)	(0.018)
Prior Suspensions (log + 1)	0.028	-0.090	-0.037	0.006	-0.017	-0.040	-0.059
	(0.029)	(0.139)	(0.163)	(0.518)	(0.476)	(0.137)	(0.129)
Prior Math Achievement	-0.004	-0.059*	-0.068**	-0.114***	-0.117***	-0.021	-0.012
	(0.009)	(0.032)	(0.027)	(0.041)	(0.038)	(0.020)	(0.022)
Prior ELA Achievement	-0.003	0.064*	0.075**	0.120***	0.122***	0.022	0.016
	(0.010)	(0.033)	(0.033)	(0.043)	(0.042)	(0.023)	(0.028)
p-value on joint test of significance	0.313	0.466	0.359	0.185	0.103	0.397	0.103
Non-Compliance/Missing Data Rates for:							
Full Sample	NA	0.312	0.296	0.207	0.218	0.421	0.483
Black Teachers	NA	0.295	0.279	0.148	0.154	0.352	0.418
White Teachers	NA	0.327	0.311	0.232	0.247	0.445	0.505
p-value on difference	NA	0.667	0.670	0.218	0.159	0.106	0.133
Students	1,283	1,283	1,283	1,283	1,283	1,283	1,283

Notes: Estimates in each column come from separate regression models that predict a dummy indicator for whether or not students' randomly assigned teacher is Black as a function of baseline student characteristics and school-grade fixed effects that are equivalent to randomization block. In column 2, baseline student characteristics are interacted with an indicator whether or not students moved classrooms after random assignment. In columns 3 through 7, baseline student characteristics are interacted with an indicator for whether or not students have outcome data on each source. To isolate comparison between the students of Black versus White teachers, models also include indicators for whether or not students' randomly assigned teacher is Asian or Hispanic (one group; estimates not shown) and interactions with baseline student characteristics. Models that assess non-compliance and missing data further condition on the interaction between these student-level indicators, the additional teacher race/ethnicity dummy, and baseline student characteristics. Joint tests of significance test the null hypothesis that baseline student characteristics (column 1) or baseline student characteristics interactd with non-compliance/missing data indicators (columns 2 through 7) are jointly equal to zero. Robust standard errors clustered at the teacher level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Supplemental Materials

Supplemental Table 1. Background Teacher Characteristics by Race/Ethnicity

	Black Teachers	Differences for:	
		White	Asian or Hispanic
Female	1.00	-0.162	-0.144
		(0.110)	(0.283)
BA in Education	0.56	0.072	-0.277
		(0.117)	(0.191)
Traditionally Certified	0.88	0.068	0.034
		(0.122)	(0.063)
MA Degree	0.75	-0.023	-0.012
		(0.185)	(0.093)
Teaching Experience (years)	10.31	-1.134	-6.005
		(1.832)	(3.629)
Math Knowledge (standardized)	-0.50	0.763*	0.773
		(0.384)	(0.607)
p-value on joint test of significance		0.463	0.520
Teachers	16	50	5

Notes: Differences in teacher background characteristics by race/ethnicity estimated from regression models that include school-grade fixed effects. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Supplemental Table 2. Item Text from Student Survey

Dimensions and Item Text
Self-Efficacy
I have pushed myself hard to completely understand math in this class.
If I need help with math, I make sure that someone gives me the help I need.
If a math problem is hard to solve, I often give up before I solve it.
Doing homework problems helps me get better at doing math.
In this class, math is too hard.
Even when math is hard, I know I can learn it.
I can do almost all the math in this class if I don't give up.
I'm certain I can master the math skills taught in this class.
When doing work for this math class, focus on learning not time work takes.
I have been able to figure out the most difficult work in this math class.
Engagement and Happiness in Class
This math class is a happy place for me to be.
Being in this math class makes me feel sad or angry. (Reverse coded.)
The things we have done in math this year are interesting.
Because of this teacher, I am learning to love math.
I enjoy math class this year.
Self-Regulation
My behavior in this class is good.
My behavior in this class sometimes annoys the teacher. (Reverse coded.)
My behavior is a problem for the teacher in this class. (Reverse coded.)

Notes: All items are on the same scale from 1 = "Strongly Disagree" to 5 = "Strongly Agree".

Supplemental Table 3. Pairwise Correlations between Student Outcomes

	Self-Efficacy	Engage.	Self-Reg.	Absences	Math Achieve.	ELA Achieve.
			Panel A: Upper-Elementary School			
Self-Efficacy	1					
Engagement and Happiness in Class	0.647***	1				
Self-Regulation	0.319***	0.241***	1			
Absences	-0.109**	-0.102**	-0.014	1		
Math Achievement	0.302***	0.215***	0.209***	-0.180***	1	
ELA Achievement	0.230***	0.124***	0.289***	-0.108***	0.643***	1
			Panel B: High School			
Self-Efficacy				-0.102***	0.259***	0.192***
Engagement/Happiness				-0.061	0.173***	0.151***
Self-Regulation				-0.010**	0.219***	0.261***
Absences				0.564***	-0.184***	-0.069*
Math Achievement				-0.174***	0.668***	0.555***
ELA Achievement				-0.153***	0.538***	0.647***

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Supplemental Table 4. Item Text from Teacher Survey

Dimensions and Item Text	Scale
Growth Mindset Beliefs	
The amount my students can learn is primarily related to family background and/or student effort.	We are interested in your ideas about intelligence. There are no right or wrong answers. Please indicate the extent to which you agree or disagree with each of the following statements. 1 = "Strongly Disagree" to 6 = "Strongly Agree" (Reverse coded)
I am limited in what I can achieve because of student home environment and/or effort.	
Students have a certain amount of intelligence, and I can't do much to change it.	
Students have certain amount of intelligence, and they can't really do much to change it.	
Intelligence is something about students that they can't change very much.	
Students can learn new things, but they can't really change their basic intelligence.	
To be honest, students can't really change how intelligent they are.	
Preparation for Instruction	
Seeking outside support for struggling students in any subject (e.g., IEPs, tutoring).	In a typical week, how much time do you devote to the following activities? 1 = "No Time" to 5 = "More than Six Hours"
Collaboratively planning lessons in any subject with other teachers or coaches.	
Grading mathematics assignments.	
Gathering and organizing mathematics lesson material (e.g., locating and copying supplemental material, preparing manipulatives).	
Reviewing the content of specific mathematics lessons (e.g., reading the teacher manual, seeking additional information about the content).	
Preparing for a mathematics lesson by trying out explanations, or working through examples of problems.	
Helping students learn any subject after school hours (e.g., homework club, tutoring).	
I differentiate mathematics assignments based on students' individualized learning needs.	About how often do you or your students take part in the following activities? 1 = "Never" to 5 = "Daily or Almost Daily"
I evaluate student work on mathematics assessments or assignments using a written rubric.	
I provide detailed written feedback on student mathematical work in addition to a numeric score.	
I examine student work to understand the process students use to solve mathematics problems.	
Students evaluate their own mathematical work on assessments or assignments using a written rubric.	
I change my lesson plans based on what I learn from analyzing student work.	
I design assignments that reveal student thinking rather than just mastery of learning goals.	
I assess students' understanding of a topic before I teach it.	
Relationships with Students and Families	
Talking with parents about students' learning or behavior.	In a typical week, how much time do you devote to the following activities? 1 = "No Time" to 5 = "More than Six Hours"
Students and I show an interest in each other's lives.	How often do you observe the following situations while teaching? 1 = "Rarely or Never" to 5 = "Always"
Students and I have a friendly rapport.	
Students and I use respectful language and listen to each other.	

Supplemental Table 5. Item Text from Classroom Assessment Scoring System (CLASS) Observation Instrument

Dimensions	Description
	Classroom Support
Positive Climate	Positive climate reflects the emotional connection and relationships among teachers and students, and the warmth, respect, and enjoyment communicated by verbal and non-verbal interactions.
Teacher Sensitivity	Teacher sensitivity reflects the teacher's timely responsiveness to the academic, social/emotional, behavioral, and developmental needs of individual students and the entire class.
Respect for Student Perspectives	Regard for student perspectives captures the degree to which the teacher's interactions with students and classroom activities place an emphasis on students' interests and ideas and encourage student responsibility and autonomy. Also considered is the extent to which content is made useful and relevant to the students.
Instructional Learning Format	Instructional learning format focuses on the ways in which the teacher maximizes student engagement in learning through clear presentation of material, active facilitation, and the provision of interesting and engaging lessons and materials.
Content Understanding	Content understanding refers to both the depth of lesson content and the approaches used to help students comprehend the framework, key ideas, and procedures in an academic discipline. At a high level, this refers to interactions among the teacher and students that lead to an integrated understanding of facts, skills, concepts, and principles.
Analysis and Problem Solving	Analysis and problem solving assesses the degree to which the teacher facilitates students' use of higher-level thinking skills, such as analysis, problem solving, reasoning, and creation through the application of knowledge and skills. Opportunities for demonstrating metacognition, i.e., thinking about thinking, are also included.
Quality of Feedback	Quality of feedback assess the degree to which feedback expands and extends learning and understanding and encourages student participation. Significant feedback may also be provided by peers.
Instructional Dialogue	Instructional dialogue captures the purposeful use of dialogue (structured, cumulative questioning and discussion which guide and prompt students) to facilitate students' understanding of content and language development.
Student Engagement	This scale captures the degree to which all students in the class are focused and participating in the learning activity presented or facilitated by the teacher.
	Classroom Organization
Negative Climate (Reverse Coded)	Negative climate reflects the overall level of negativity among teachers and students in the class.
Behavior Management	Behavior management encompasses the teacher's use of effective methods to encourage desirable behavior and prevent and redirect misbehavior.
Productivity	Productivity considers how well the teacher manages time and routines so that instructional time is maximized. This dimension captures to degree to which instructional time is effectively managed and down time is minimized for students.

Notes: Item text comes directly from CLASS upper-elementary manual (Pianta, Hamre, & Mintz, 2012).

Supplemental Table 6. Pairwise Correlations between Teacher Mindsets and Practices

	Growth Mindset Beliefs	Preparation for Instruction	Relationships with Students and Families	Classroom Support	Classroom Organization
Growth Mindset Beliefs	1				
Preparation for Instruction	0.227*	1			
Relationships with Students and Families	0.050	0.432***	1		
Classroom Support	0.120	0.373***	0.222*	1	
Classroom Organization	0.290**	0.322***	0.017	0.302**	1

*** p<0.01, ** p<0.05, * p<0.1

Supplemental Table 7. Average Effects of Asian or Hispanic Teachers versus White Teachers on Student Outcomes

Student Outcomes		(1)	(2)	(3)	(4)
		Panel A: End-of-Year Effects			
Self-Efficacy	903	0.436**	0.379**	0.461***	0.404**
		(0.181)	(0.187)	(0.170)	(0.183)
Engagement/Happiness	903	0.208	0.135	0.252	0.177
		(0.194)	(0.179)	(0.175)	(0.171)
Self-Regulation	901	-0.013	-0.102	-0.044	-0.131
		(0.126)	(0.130)	(0.120)	(0.123)
Absences (log + 1)	1,017	-0.059	-0.155***	-0.079	-0.180***
		(0.133)	(0.058)	(0.113)	(0.061)
Math Achievement	1,003	0.581***	0.394***	0.585***	0.413***
		(0.119)	(0.087)	(0.097)	(0.080)
ELA Achievement	1,004	0.381***	0.260**	0.368***	0.277***
		(0.140)	(0.110)	(0.123)	(0.099)
		Panel B: Follow-Up Effects in High School			
Absences (log + 1)	743	-0.120	-0.182**	-0.180*	-0.246***
		(0.162)	(0.078)	(0.101)	(0.062)
Math Achievement	663	0.300*	0.157	0.321*	0.152
		(0.176)	(0.118)	(0.176)	(0.127)
ELA Achievement	670	0.327***	0.176	0.354***	0.217*
		(0.098)	(0.114)	(0.115)	(0.124)
School-Grade Fixed Effects		X	X	X	X
Background Student Characteristics			X		X
Background Teacher Characteristics				X	X

Notes: Estimates in each cell come from a separate regression model that predicts the outcome listed in each row on dummy indicators or whether or not students' randomly assigned teacher was Asian or Hispanic, Black (estimates not shown; see main text), with White as the left-out category; and school-grade fixed effects. Some models include background teacher characteristics: gender, BA in education, traditional certification, MA degree, years of teaching experience, and math knowledge. Some models include background student characteristics include: gender, race/ethnicity, eligibility for free or reduced-price lunch, eligibility to receive special education services, limited English proficiency status, and prior-year absences, suspensions, and achievement in math and ELA. All outcomes other than absences are standardized. Robust standard errors clustered at the teacher level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Supplemental Table 8. Average Effects of Black versus White Teachers on Student Outcomes, Multiply Imputing Missing Student Outcomes

Student Outcome	Sample	Average Effects Across All Students				Subgroup Effects	
						Black Students	Non-Black
		(1)	(2)	(3)	(4)	(5)	(6)
Panel A: End-of-Year Effects in Upper-Elementary School							
Self-Efficacy	1,283	0.350***	0.351***	0.395***	0.398***	0.684***	0.229
		(0.132)	(0.126)	(0.131)	(0.125)	(0.211)	(0.177)
Engagement/Happiness	1,283	0.274**	0.249*	0.249*	0.227	0.552***	0.018
		(0.132)	(0.126)	(0.148)	(0.144)	(0.202)	(0.184)
Self-Regulation	1,283	-0.020	-0.036	-0.016	-0.009	0.030	-0.002
		(0.116)	(0.111)	(0.122)	(0.117)	(0.225)	(0.155)
Absences (log + 1)	1,283	-0.189***	-0.222***	-0.238***	-0.252***	-0.325**	-0.227**
		(0.066)	(0.057)	(0.072)	(0.061)	(0.139)	(0.114)
Math Achievement	1,283	0.150*	0.158***	0.208**	0.228***	0.122	0.234**
		(0.077)	(0.057)	(0.084)	(0.065)	(0.129)	(0.102)
ELA Achievement	1,283	0.098	0.124*	0.128	0.196***	0.264**	0.147
		(0.094)	(0.065)	(0.089)	(0.063)	(0.129)	(0.112)
Panel B: Follow-Up Effects in High School							
Absences (log + 1)	1,283	-0.087	-0.105	-0.099	-0.100	-0.258	-0.024
		(0.090)	(0.084)	(0.101)	(0.096)	(0.192)	(0.173)
Math Achievement	1,283	0.018	0.004	0.002	-0.002	0.006	-0.020
		(0.093)	(0.076)	(0.085)	(0.066)	(0.153)	(0.112)
ELA Achievement	1,283	0.142	0.142	0.100	0.132	0.235	0.076
		(0.102)	(0.092)	(0.108)	(0.094)	(0.192)	(0.134)
School-Grade Fixed Effects		X	X	X	X	X	X
Background Student Characteristics			X		X	X	X
Background Teacher Characteristics				X	X	X	X

Notes: Estimates in each cell come from a separate regression model that regresses the outcome listed in each row on dummy indicators or whether or not students' randomly assigned teacher was Black, Asian or Hispanic (not shown), with White as the left-out category; and school-grade fixed effects. Some models include background teacher characteristics: gender, BA in education, traditional certification, MA degree, years of teaching experience, and math knowledge. Some models include background student characteristics include: gender, race/ethnicity, eligibility for free or reduced-price lunch, eligibility to receive special education services, limited English proficiency status, and prior-year absences, suspensions, and achievement in math and ELA. All outcomes other than absences are standardized. Robust standard errors clustered at the teacher level in parentheses. Parameters are estimated across 10 replication data sets that multiply impute missing student outcomes as a function of baseline student and teacher characteristics, and fixed effects for randomization block. *** p<0.01, ** p<0.05, * p<0.1

Supplemental Table 9. Effects of Black versus White Teachers on Student Outcomes After Accounting for Student-Level Mediators

Student Outcome	Student Group	Sample	Total Effect	Direct Effect After Accounting for:					
				Self-Effiacy	Engage./ Happiness	Self-Regulation	Absences	Math Achieve.	ELA Achieve.
Panel A: End-of-Year Effects in Upper-Elementary School									
Self-Efficacy	Black	356	0.853***	--	0.245***	0.389***	--	--	--
			(0.217)		(0.077)	(0.115)			
Engagement/Happiness	Black	356	0.689***	0.129	--	0.620***	--	--	--
			(0.171)	(0.100)		(0.153)			
Self-Regulation	Black	356	0.216	-0.026	-0.003	--	--	--	--
			(0.169)	(0.159)	(0.150)				
Absences (log + 1)	Black	356	-0.465***	-0.404***	-0.421***	-0.462***	--	--	--
			(0.101)	(0.110)	(0.106)	(0.103)			
Absences (log + 1)	Non-Black	545	-0.172***	-0.172***	-0.173***	-0.173***	--	--	--
			(0.056)	(0.057)	(0.056)	(0.058)			
Math Achievement	Black	355	0.081	-0.002	-0.005	0.066	0.023	--	--
			(0.140)	(0.135)	(0.129)	(0.133)	(0.144)		
Math Achievement	Non-Black	545	0.279***	0.271***	0.283***	0.280***	0.265***	--	--
			(0.072)	(0.070)	(0.071)	(0.072)	(0.071)		
ELA Achievement	Black	356	0.342**	0.316**	0.329**	0.321**	0.314**	--	--
			(0.149)	(0.150)	(0.146)	(0.151)	(0.154)		
Panel B: Follow-Up Effects in High School									
Absences (log + 1)	Black	245	-0.536***	-0.442***	-0.450***	-0.513***	-0.328**	-0.506***	-0.502***
			(0.130)	(0.150)	(0.146)	(0.138)	(0.143)	(0.138)	(0.123)
Math Achievement	Black	223	0.170	0.072	0.096	0.189	0.070	0.110	0.118
			(0.141)	(0.140)	(0.141)	(0.135)	(0.159)	(0.108)	(0.125)
ELA Achievement	Black	226	0.163	0.093	0.085	0.159	0.127	0.121	0.082
			(0.164)	(0.160)	(0.157)	(0.161)	(0.168)	(0.131)	(0.134)
ELA Achievement	Non-Black	370	0.063	0.063	0.052	0.059	0.054	0.018	0.019
			(0.124)	(0.124)	(0.125)	(0.118)	(0.122)	(0.119)	(0.117)

Notes: Estimates in each cell estimate the effect of Black versus White teachers on the student outcome listed in column 1, after accounting for the hypothesized mediator listed in column headers. Estimates come from the preferred model with background student and teacher characteristics, as well as school-grade fixed effects. Sample sizes differ from the main results, as the mediation analyses include a consistent sample of students who have data on each outcome and all relevant mediators. "--" indicates the student outcome listed in each column is not used as a medaitor when predicting the outcomes in each row. *** p<0.01, ** p<0.05, * p<0.1

Supplemental Table 10. Effects of Black versus White Teachers on Student Outcomes After Accounting for Teacher-Level Mediators

Student Outcome	Student Group	Sample	Total Effect	Direct Effect After Accounting for:				
				Growth Mindset	Prep. for Inst.	Relat. Stu./Fam.	Class Support	Class Org.
Panel A: End-of-Year Effects in Upper-Elementary School								
Self-Efficacy	Black	357	0.853***	0.984***	0.496***	0.807***	1.152***	0.451*
			(0.217)	(0.212)	(0.180)	(0.208)	(0.296)	(0.241)
Engagement/Happiness	Black	357	0.685***	0.765***	0.509***	0.682***	0.768***	0.421
			(0.172)	(0.147)	(0.171)	(0.173)	(0.218)	(0.256)
Self-Regulation	Black	356	0.216	0.202	0.015	0.251	0.506***	0.120
			(0.169)	(0.185)	(0.206)	(0.170)	(0.177)	(0.221)
Absences (log + 1)	Black	407	-0.387***	-0.403***	-0.346***	-0.390***	-0.360***	-0.585***
			(0.106)	(0.123)	(0.128)	(0.105)	(0.110)	(0.138)
Absences (log + 1)	Non-Black	610	-0.196***	-0.145***	-0.204***	-0.150***	-0.118	-0.282***
			(0.043)	(0.048)	(0.049)	(0.056)	(0.073)	(0.058)
Math Achievement	Black	401	0.139	0.189	0.106	0.148	0.256**	-0.031
			(0.134)	(0.163)	(0.121)	(0.124)	(0.108)	(0.173)
Math Achievement	Non-Black	602	0.202***	0.134	0.129	0.263***	0.259***	0.084
			(0.074)	(0.111)	(0.080)	(0.089)	(0.088)	(0.084)
ELA Achievement	Black	402	0.387***	0.236	0.410***	0.385***	0.420*	0.341**
			(0.141)	(0.151)	(0.089)	(0.140)	(0.226)	(0.134)
Panel B: Follow-Up Effects in High School								
Absences (log + 1)	Black	283	-0.355***	-0.582***	-0.478***	-0.424***	-0.533***	-0.274**
			(0.128)	(0.161)	(0.146)	(0.122)	(0.190)	(0.129)
Math Achievement	Black	260	0.115	0.101	0.163	0.174	0.228	-0.120
			(0.117)	(0.153)	(0.153)	(0.133)	(0.198)	(0.125)
ELA Achievement	Black	262	0.174	0.047	0.261*	0.194	0.094	0.135
			(0.125)	(0.116)	(0.133)	(0.119)	(0.177)	(0.126)
ELA Achievement	Non-Black	408	0.216*	0.416***	0.167	0.233**	0.284*	0.244
			(0.115)	(0.122)	(0.128)	(0.111)	(0.160)	(0.184)

Notes: Estimates in each cell estimate the effect of Black versus White teachers on the student outcome listed in column 1, after accounting for the hypothesized mediator listed in column headers. Estimates come from the preferred model with background student and teacher characteristics, as well as school-grade fixed effects. *** p<0.01, ** p<0.05, * p<0.1

Supplemental Table 11. Relationships between Teacher-Level Mediators and Student Outcomes

	Upper-Elementary School							High School			
Teacher Mindset or Practice	Self-Efficacy	Engage./ Hap.	Absences (log + 1)	Absences (log + 1)	Math Achieve.	Math Achieve.	ELA Achieve.	Absences (log + 1)	Math Achieve.	ELA Achieve.	ELA Achieve.
Growth Mindset Beliefs	-0.243**	-0.205**	-0.035	0.048	-0.050	-0.083	0.120	0.046	-0.165*	-0.102	-0.189***
	(0.116)	(0.092)	(0.053)	(0.040)	(0.084)	(0.068)	(0.077)	(0.099)	(0.085)	(0.089)	(0.052)
Growth Mindset Beliefs*Black Teacher	-0.007	0.079	0.008	-0.085	-0.007	0.137	0.029	0.279***	0.161	0.287***	0.140
	(0.120)	(0.123)	(0.074)	(0.054)	(0.095)	(0.098)	(0.087)	(0.082)	(0.103)	(0.108)	(0.089)
Prep. for Instruction	0.212	0.249	-0.007	-0.095**	-0.086	-0.055	-0.129	0.524***	-0.363***	-0.140	0.127
	(0.136)	(0.152)	(0.088)	(0.048)	(0.113)	(0.096)	(0.096)	(0.129)	(0.127)	(0.116)	(0.078)
Prep. for Instruction*Black Teacher	0.425	0.062	-0.124	0.095	0.150	0.128	0.031	-0.418**	0.284	-0.140	-0.024
	(0.264)	(0.217)	(0.110)	(0.067)	(0.111)	(0.115)	(0.170)	(0.157)	(0.208)	(0.182)	(0.179)
Relationships with Stu./Fam.	-0.126	-0.004	-0.024	0.101*	-0.114	0.198**	-0.052	0.162	-0.106	0.039	0.178**
	(0.154)	(0.162)	(0.078)	(0.054)	(0.113)	(0.078)	(0.087)	(0.141)	(0.148)	(0.098)	(0.072)
Relationships with Stu./Fam.*Black Teacher	0.218	-0.019	0.012	-0.068	0.135	-0.230***	-0.115	0.046	-0.068	-0.096	0.032
	(0.227)	(0.219)	(0.116)	(0.055)	(0.155)	(0.084)	(0.136)	(0.160)	(0.153)	(0.107)	(0.085)
Classroom Support	0.113	0.313***	-0.033	-0.032	0.123***	0.162***	0.113**	0.088	-0.057	-0.022	0.027
	(0.076)	(0.060)	(0.073)	(0.029)	(0.046)	(0.043)	(0.047)	(0.070)	(0.130)	(0.073)	(0.065)
Classroom Support*Black Teacher	-0.607**	-0.521***	-0.013	-0.069	-0.307***	-0.116	-0.193	0.131	-0.051	0.125	-0.090
	(0.250)	(0.187)	(0.109)	(0.068)	(0.091)	(0.097)	(0.175)	(0.148)	(0.201)	(0.138)	(0.159)
Classroom Organization	-0.121	0.008	0.055	0.081**	0.102	0.198***	0.149**	0.014	0.203**	0.138	0.016
	(0.132)	(0.131)	(0.075)	(0.040)	(0.062)	(0.043)	(0.065)	(0.084)	(0.079)	(0.086)	(0.073)
Classroom Organization*Black Teacher	0.891**	0.461	0.296*	0.102	0.171	0.007	-0.179	-0.226	0.146	-0.098	-0.080
	(0.380)	(0.309)	(0.167)	(0.088)	(0.110)	(0.099)	(0.243)	(0.210)	(0.123)	(0.177)	(0.267)
Student Group	Black	Black	Black	Non-Black	Black	Non-Black	Black	Black	Black	Black	Non-Black
Students	357	357	407	610	401	602	402	283	260	262	408

Notes: Estimates in each column and pair of two rows come from a separate regression model that regresses the outcome on dummy indicators or whether or not students' randomly assigned teacher was Black, Asian or Hispanic, with White as the left-out category; mediators, and their interactions with dummy indicators for teacher being Black (shown), Asian or Hispanic (not shown); student background characteristics; and school-grade fixed effects. All outcomes other than absences are standardized. Robust standard errors clustered at the teacher level in parentheses. *** p<0.01, ** p<0.05, * p<0.01

Supplemental Table 12. Tests of Sequential Ignorability: Pairwise Correlations between Residuals from Model that Regresses Mediator on Treatment and Model that Regresses Outcome on Treatment and Mediator

Outcome	Student Group	Student-Level Mediators						Teacher-Level Mediators				
		Self-Effiacy	Engage./ Happiness	Self-Regulation	Absences	Math Achieve.	ELA Achieve.	Growth Mindset	Prep. for Inst.	Relat. Stu./Fam.	Class Support	Class Org.
Panel A: End-of-Year Effects in Upper-Elementary School												
Self-Efficacy	Black	--	-0.031	0.021	--	--	--	0.000	0.000	0.000	0.000	0.000
Engagement/Happiness	Black	-0.003	--	-0.091	--	--	--	0.000	0.000	0.000	0.000	0.000
Self-Regulation	Black	0.033	-0.088	--	--	--	--	0.000	0.000	0.000	0.000	0.000
Absences (log + 1)	Black	0.018	0.000	-0.029	--	--	--	0.000	0.000	0.000	0.000	0.000
Absences (log + 1)	Non-Black	-0.056	-0.057	0.025	--	--	--	0.000	0.000	0.000	0.000	0.000
Math Achievement	Black	0.033	-0.044	0.050	0.013	--	--	0.000	0.000	0.000	0.000	0.000
Math Achievement	Non-Black	0.005	0.055	0.028	-0.013	--	--	0.000	0.000	0.000	0.000	0.000
ELA Achievement	Black	0.123	0.113	0.070	0.015	--	--	0.000	0.000	0.000	0.000	0.000
Panel B: Follow-Up Effects in High School												
Absences (log + 1)	Black	0.059	0.056	-0.017	0.006	0.091	-0.013	0.000	0.000	0.000	0.000	0.000
Math Achievement	Black	-0.076	-0.108	-0.067	0.030	-0.148	0.052	0.000	0.000	0.000	0.000	0.000
ELA Achievement	Black	-0.009	-0.068	-0.048	0.028	-0.202	-0.127	0.000	0.000	0.000	0.000	0.000
ELA Achievement	Non-Black	0.029	0.071	-0.016	-0.002	0.032	0.004	0.000	0.000	0.000	0.000	0.000

Notes: "--" indicates the student outcome listed in each column is not used as a medaitor when predicting the outcomes in each row.