ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 21-501

How and Why Do Black Teachers Benefit Students?: An Experimental Analysis of Causal Mediation

David Blazar
University of Maryland College Park

Using a causal mediation framework, I find several social dynamics that explain how and why Black teachers benefit students. Random assignment to a Black versus a White teacher in upper-elementary school increases self-efficacy and engagement of Black students (0.6 SD), and increases test scores (0.2 SD) and decreases chronic absenteeism (60% reduction) of all students. These total effects are partly explained by "good" teaching practices and mindsets that Black teachers possess more than White teachers. However, the measures do not fully mediate the total effects of Black teachers, indicating that other social interactions such as role modeling also play a role. The findings provide motivation for recruiting more Black teachers and insight into training the current, mostly White teacher workforce.

VERSION: August 2022

Suggested citation: Blazar, David. (2022). How and Why Do Black Teachers Benefit Students?: An Experimental Analysis of Causal Mediation. (EdWorkingPaper: 21-501). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/-jym0-wz02

How and Why Do Black Teachers Benefit Students?: An Experimental Analysis of Causal Mediation

David Blazar
University of Maryland

August 2022

Abstract

Using a causal mediation framework, I find several social dynamics that explain how and why Black teachers benefit students. Random assignment to a Black versus a White teacher in upper-elementary school increases self-efficacy and engagement of Black students (0.6 SD), and increases test scores (0.2 SD) and decreases chronic absenteeism (60% reduction) of all students. These total effects are partly explained by "good" teaching practices and mindsets that Black teachers possess more than White teachers. However, the measures do not fully mediate the total effects of Black teachers, indicating that other social interactions such as role modeling also play a role. The findings provide motivation for recruiting more Black teachers and insight into training the current, mostly White teacher workforce.

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

Correspondence regarding the paper can be sent to David Blazar at dblazar@umd.edu; Department of Teaching and Learning, Policy and Leadership, 2311 Benjamin Building, University of Maryland College of Education, College Park, MD, 20742. I thank staff in the research offices of collaborating districts that provided data; Casey Archer at the University of Maryland for research support; and Seth Gershenson, Ramon Goings, David Liebowitz, Susanna Loeb, and anonymous reviewers for providing feedback on earlier versions of the paper.

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; Jencks & Phillips, 2011; Reardon & Galindo, 2009), high school graduation (Hernandez, 2011; Murnane, 2013), college going (Merolla, 2018), 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 (Irvine, 1989; Graham, 1987; Ladson-Billings, 1994; Waters, 1989). Building from this theory, causally oriented studies document substantively meaningful teacher-student race/ethnicity-matching effects on students' academic outcomes including test-based achievement (for reviews, see Bristol & Martin-Fernandez, 2019; Redding, 2019), absences and suspensions from school (Gottfried et al., 2021; Holt & Gershenson, 2019; Lindsay & Hart, 2017; Shirrell et al., 2021; Tran & Gershenson, 2021), and educational attainment (Delhommer, 2022). For example, in the only other experimental dataset—aside from the current study—used to examine the impact of Black teachers[1], Dee (2004) estimated effects on the end-of-year test scores of Black students of more than

[1] Egalite and Kisida (2018) estimated race/ethnicity-matching effects in data from the Measures of Effective Teaching (MET) project, where a subsample of teachers was randomly assigned to class rosters within schools. However, the authors relied primarily on the observational/non-experimental portion of the dataset, noting that "noncompliance within the

0.20 standard deviations (SD), representing roughly one-third of the Black-White test score gap (Fryer & Levitt, 2004; Jencks & Phillips, 2011). In the same sample, these short-term effects translated into key markers of human capital development for Black students, including increased rates of high school graduation and college going (4 to 5 percentage points; Gershenson et al., 2021).

The current analyses build from the prior literature by examining the underlying mechanisms driving the effects of Black teachers on student outcomes. That is: *How* and *why* is that assignment to a Black teacher benefits student outcomes? Orienting the quantitative teacher-student race/ethnicity-matching literature towards "how" and "why" questions is critical for at least two reasons. First, the theoretical literature 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 by exemplifying academic success, which in turn drives improved outcomes (Fordham & Ogbu, 1986)? Do Black teachers also engage in "culturally relevant" (Ladson-Billings, 1995b), "culturally responsive" (Gay, 2002), and "culturally sustaining" (Paris, 2012) pedagogies that are particularly beneficial for their students of color—or that may benefit all students? In his discussion of race/ethnicity-based opportunity and achievement gaps, Ferguson (2003) describes how teachers' perceptions, expectations, and practices shape classroom interactions and likely are reflected in students' own perceptions, expectations, behaviors, and outcomes. In other words, the path between Black teachers and the outcomes of their students almost certainly run through a set of classroom social dynamics.

Yet, the quantitative, causal literature on the social interactions of Black teachers with their students is quite limited. Most studies of teacher-student race/ethnicity matching focus on effects on

randomization procedure…tempers our ability to be absolutely certain that the effects we identify are causal" (p. 64). Further, Constantine et al. (2009) reported positive effects of Black teachers on the test scores of Black students using another dataset in which teachers were randomly assigned to classes. However, the primary focus of this study was teachers' certification pathway and not teacher-student race-matching effects. As such, the latter estimates were reported as supplemental analyses and without specific attention paid to causal assumptions related to this question.

student test scores (Redding, 2019), which provides an important but incomplete picture of student development. Scholars have long advocated for increased access to Black teachers in order to improve the academic achievement of Black students (Ladson-Billings, 1994; Ferguson, 2003), while also noting that test scores are most likely to improve when classroom conditions are conducive to learning (e.g., students feel engaged and empowered; Gay, 2000; Milner, 2011). A handful of recent studies document benefits of race/ethnicity matching to student engagement in learning activities, motivation, and social skills (Egalite & Kisida, 2018; Rasheed et al., 2020; Wright et al., 2017). But, none of these studies are experimental, and they do not examine teacher-level perceptions, expectations, and practices that may drive these effects. Comparatively, a couple of studies document how teacher-student race/ethnicity matching reduces teachers' perceptions that Black students are disruptive or inattentive (Dee, 2005), as well as biases in teachers' expectations for students' educational attainment (Gershenson et al., 2016), However, these studies also are not experiments, and they do they link to student-level outcomes.

A second reason to probe underlying mechanisms of teacher-student race/ethnicity-matching effects is that stakeholder action depends on knowing the answer. As Gershenson et al. (2021) point out, if the effects of Black teachers on the outcomes of their students are explained by a set of mindsets, practices, and skills that they bring to their work, then there may be opportunities through pre- or in-service development to train the mostly-White teacher workforce in these areas. Alternatively, if the effects are driven primarily or exclusively 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.

To bridge these gaps in the literature, I draw on a unique dataset where fourth- and fifth-grade teachers were randomly assigned to class rosters within schools, paired with rich data on varied student outcomes and varied teacher mindsets and practices. With these data, I first estimate the impacts of

Black versus White teachers on intrapersonal components of students' social-emotional development (i.e., self-efficacy, engagement, self-regulation), test scores in math and English language arts (ELA), and observed school behaviors (i.e., absences, suspensions), captured both at the end of the year working with that teacher and several years later when students are in high school. I estimate effects of Black teachers on these outcomes on average across all students, as well as race-matching effects for Black students specifically. Distinguishing between average and race-matching effects can 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.

Next, I examine whether there are differences between Black and White teachers in their mindsets regarding student learning and their instructional practices, and whether these differences mediate average effects on student outcomes. I focus on teacher mindset and practice measures that fall under what Ladson-Billings (1995a) refers to as "good" teaching practices in her discussion of culturally relevant pedagogy, and that are documented qualitatively in the classrooms of Black teachers (e.g., Milner, 2011; Ware, 2006). These practices include teachers' beliefs that student intelligence is malleable rather than fixed, their relationships with students and families, their preparation for and differentiation of instruction, and their ability to lead well-organized classrooms without creating a negative classroom climate. That said, the available data do not include all measures of culturally relevant or culturally responsive teaching (e.g., teachers' support for "cultural competence" and "critical consciousness"), and so this study should be interpreted as an impact evaluation of these pedagogies as a whole.

Using a causal mediation framework to link teachers' race, their mindsets and practices, and varied student outcomes (Baron & Kenny, 1986; Imai et al., 2010; VanderWeele, 2015), I find a multidimensional landscape of social dynamics that help explain how and why Black teachers benefit their students' outcomes. First, random assignment to a Black teacher in upper-elementary school

results in very large total effects on the intrapersonal competencies of their Black students (0.56 standard deviations [SD] for self-efficacy, and 0.66 SD for engagement). I also find impacts of Black teachers on test scores (0.12 SD in math and 0.2 SD in ELA) and absences (roughly 60% reduction in chronic absenteeism relative to the rate for students of White teachers) that are similar to other experimental studies (Dee, 2004; Tran & Gershenson, 2021). These total effects often extend to all students no matter their race or ethnicity, and they also persist at similar magnitudes up to six years later when students are in high school.

Several of these total effects are explained in part by a set of "good" teaching practices and mindsets that Black teachers tend to possess more than White teachers (between-group differences upwards of 0.3 SD). Black teachers' ability to lead well-organized classrooms in which student (mis)behavior is addressed productively without creating a negative classroom climate partially explains their effects on students' end-of-year math achievement (roughly 25% of the total effect) and student-reported self-regulation (roughly 50%). Black teachers also are more likely than White teachers to hold growth mindset beliefs that student intelligence is malleable rather than fixed, which partially explains their effects on end-of-year student engagement (roughly 7%) and longer-run chronic absenteeism (roughly 12%). At the same time, the set of available teacher mindset and practice measures do not fully mediate the effects of Black versus White teachers on any outcome, indicating that other social interactions also are at play. For example, I find some of the largest total effects of Black teachers on the self-efficacy of their Black students, but I do not find any mediating pathways here. Given the nature of the self-efficacy as a construct related to confidence and self-image, it may be that Black teachers' position as role models for their Black students drives these effects.

Together, the findings provide a rich picture of the role that Black teachers play in supporting their students, as well as avenues for intervention. To address the needs of the current—mostly White—teacher workforce, training and professional development may focus specifically on the

mindsets and practices that drive improved outcomes (i.e., classroom organizational skills, growth mindset beliefs). Training efforts must also be paired by widescale recruitment of Black and other prospective teachers of color aimed at dramatically shifting demographics of the current educator workforce. This study shows that Black students benefit from having role models who look like them in positions of power, and that non-Black students benefit from work with Black teachers too. Further, I present evidence of exclusionary discipline of Black students assigned to a White teacher (3.7 percentage point differential in suspension rates, compared to White students), which is unlikely to be dismantled through training alone. Moving forward, policy and practice action must recognize the incredible resources that Black teachers provide to students, and that we need to support many more of them to enter and stay working in schools.

Motivating Literature

Although the research literature linking teacher-level characteristics to student outcomes has crystallized around the benefit of same-race/ethnicity matching (Bristol & Martin-Fernandez, 2019; Redding, 2019), 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 (Fordham & Ogbu, 1986). Increased diversity amongst schools' professional staff can help offset the normalization of race/ethnicity-based stratification that students of color experience in their lives both inside and outside of school (Villegas & Lucas, 2004).

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. Scholars describe how the academic experiences and outcomes of students of color are informed by their lives beyond the

classroom. As such, it is important that they have teachers who recognize and seek to understand how racial inequality shapes their world (Irvine, 1989; Ladson-Billings, 1994). White teachers are not inherently unable to teach students of color, but may be more likely than teachers of color 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; Valencia, 1997). Compared to White teachers, teachers of color may also have higher expectations for students of color (Ferguson, 2003), which can be critical for offsetting "stereotype threat" and the risk of confirming a negative stereotype about a group (Steele & Aronson, 1995).

Third, Black and other teachers of color may be better at teaching *all* students. Whether Black, Hispanic, Asian, or White, students report feeling better cared for and more academically challenged when they have a teacher of color (Cherng & Halpin, 2016). In other words, the practices and behaviors that teachers of color may deliver with students of color in mind may just be "good teaching" all around (Ladson-Billings, 1995a). There is a large literature, for example, describing how teachers' interpersonal relationships with students benefits an array of academic and social-emotional outcomes, on average across teachers and students from different backgrounds (Perlam et al., 2016; Pianta & Hamre, 2009).

However, scholars generally have been quite limited by available data to explore these mechanisms and mediating pathways in any rigorous way. The current study builds most directly from Gershenson et al.'s (2021) work, which exploits the random assignment of teachers to students in the Project STAR experiment from 1908s Tennessee to identify race/ethnicity-matching effects on short- and long-run student outcomes, as well as to explore possible mechanisms underlying these effects. The authors found that the effects of Black teachers on short-term test scores and longer-run outcomes at the end of high school extended only to Black students and not to White students, which they interpreted as evidence that Black teachers are not necessarily more effective overall and likely

serve as role models specifically to their Black students. In further support of this claim, they found Black teacher-student matching effects were similar when including/excluding observable background characteristics of teachers (i.e., teaching experience, highest degree attained, status on a career ladder). Yet, the authors also acknowledged that identifying exact mechanisms through which the effect of same-race/ethnicity teachers runs requires additional data. The background characteristics of teachers in the Project STAR dataset are indirect proxies for effective teaching (Stronge, 2018). And, they do not align with the theoretical literature on teaching practice and pedagogy that emphasizes more specific channels.

In making her case for culturally relevant pedagogy, Ladson-Billings (1995a) argued that the "pedagogical excellence" of Black teachers includes key features of "good teaching" described across the teaching literature (e.g., pedagogical content knowledge); but it also is more than that (p. 159). First and foremost, "good" teach*ers* and "good" teach*ing* must support their students—and their Black students in particular—to achieve academic success. Low expectations for Black students create hostile classroom environments that are not conducive to learning, let alone overall student development. Ladson-Billings (1995a) wrote that, "…the need to improve the self-esteem of African-American students" demands teachers to situate high expectations for academic success at the "base" of instruction (p. 160). Support for academic excellence can emerge through a combination of teacher mindsets and expectations that all students can learn, in- and out-of-school efforts to build strong interpersonal relationships with students to support engagement in the classroom environment, and differentiating instruction to meet the needs of individual students as they pursue academic goals (Gay, 2000; Ladson-Billings, 1995b; Paris, 2012). The culturally relevant pedagogy of Black teachers also is described as one of "opposition", where teachers use their understanding of students' culture and lives outside of the classroom to guide instruction, as well as support for students to critique cultural norms, values, and institutions that produce and maintain social inequities.

Over several decades, scholars have engaged in qualitative exploration of classrooms, providing rich descriptions of how Black teachers (and others) enact culturally relevant, culturally responsive, and culturally sustaining pedagogies. This line of inquiry suggests that students in these environments are more engaged than students in classrooms without these features (e.g., Ladson-Billings, 1995b; Milner, 2011; Ware, 2006). More recently, research teams have developed observation instruments to identify and score the quality of teachers' classroom practice along different dimensions of practice described above (e.g., Goffney, 2010; Jensen et al., 2018; Powell et al., 2016). However, few studies have linked these observed classroom and teaching practices to student outcomes. One exception is a study by Larson et al. (2018), which found a positive association between positive student behavior and a measure that asked classroom observers to look for instances of strong teacher-student interactions, connections between content and real-world examples, integration of cultural artifacts into learning activities, and others. The authors noted that, like the broader literature base, their study is limited in its ability to draw causal conclusions because neither teachers nor the practices were randomly assigned to classes or students.

There is slightly more literature that focuses on race/ethnicity-based differences in teachers' beliefs and expectations for students' academic performance and attainment. In addition to studies described briefly in the Introduction (Dee, 2005; Gershenson et al., 2016), Papageorge et al. (2020) found racial biases in teachers' expectations for students' long-term educational attainment. Exploiting non-experimental variation within student and across teachers, the authors found that White teachers were substantially more optimistic about White students' ability to complete a four-year college degree compared to their beliefs on Black students' educational attainment. On average across teachers from different backgrounds, teachers' academic expectations for students affected actual college degree attainment. However, the authors did not examine whether differences in academic expectations between teachers of color and White teachers mediated effects on longer-run student outcomes.

There also is some relevant discussion of the effect of instructional programs (e.g., professional development, curricula) aimed at supporting teachers to develop the "good" practices and mindsets descried above, as well subsequent effects on students. Here too, though, the evidence base is slim. In a recent comprehensive review, Bottiani et al. (2018) identified just two quantitative studies that compared outcomes of participants in culturally responsive teaching interventions versus those not exposed to the interventions. One study focused on diversity training (Thompson & Byrnes, 2011), while the other focused on a schoolwide program that used data-based decision making and support from school leaders to help teachers create classroom environments with a focus on culture and racial/ethnic equity (Vincent et al., 2011). However, both of these studies failed to meet evidence standards for supporting causal inferences. Since that review, Dee and Penner (2017) found positive effects on student outcomes of an ethnic studies curriculum in California with features aligned to culturally responsive teaching, as well as an instructional and mentoring program for Black males developed through the Obama administration's My Brother's Keeper program (Dee & Penner, 2019). Both used causally-oriented research designs, but were not randomized control trials.

Ultimately, scholars have called for more evidence—and more rigorous, causal evidence in particular—on the mechanisms linking Black and other teachers of color to improved student outcomes, and the extent to which "good" teaching practices and mindsets aligned to culturally relevant pedagogy and culturally responsive teaching serve as key mediators (Bottiani et al., 2018; Gershenson et al., 2021; Hill, 2020). The current study aims to fill this gap.

Sample and Experimental Design

The data used in 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). A primary content focus of the study was mathematics, though participating teachers were generalists who taught all core subjects,

and data collection efforts and instruments crossed content areas. Over three school years (2010-11 through 2012-13), the research team collaborated with teachers in four school districts on the east coast of the U.S. to collect a variety of teacher-level measures through a set of teacher surveys and observations of classroom, and then to link these measures to researcher-collected student surveys and district-collected test scores, absences, and suspension records (see below for in-depth discussion of the teacher and student measures). The current analyses also link the project-organized data collection with administrative records on student test scores, absences, and suspensions provided by the partner districts through the 2018-19 school year (i.e., the last school year before the Covid-19 pandemic interrupted districts' collection of these measures).

During the third year of the study, a subset of teachers ($n = 71$) agreed to be randomly assigned to class rosters within schools. In spring 2012, the project team worked with staff at participating schools to randomly assign sets of teachers to class rosters ($n = 1{,}283$ students) of the same grade level that were constructed by principals or other school leaders. To be eligible for randomization teachers had to work in schools and grades in which there was at least one other participating teacher. Their principal also had to consider these teachers as capable of teaching any of the rosters of students designated for the group of teachers.[2]

The project's volunteer sample matches the characteristics of teachers and students across the four participating school districts, and of urban school districts in the U.S. more broadly (see Table 1; U.S. Department of Education, 2019). The vast majority of participating teachers were female (85%) and White (70%), with roughly 11 years of teaching experience. Of the participating teachers of color,

[2] 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 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.

the vast majority were Black (23%); a small subset of teachers was Asian (4%) or Hispanic (3%). Driven largely by sample size restrictions, the analyses presented in this paper compare Black and White teachers only. Black and White teachers in the experimental sample were similar to each other in terms of degrees (75% and 79% of Black and White teachers, respectively, have master's degrees), certification pathway (88% and 93% were traditionally certified), and teaching experience (10.3 years and 11.7 years), with no statistically significant differences. This is important, as it means that other differences between Black and White teachers on their mindsets and practices are not simultaneously capturing differences in training and improvement over time. All of the Black teachers in the sample were female compared to 80% of the White teachers, which is consistent with national patterns showing that Black males are substantially underrepresented in the teacher workforce (Britsol & Goings, 2019). Compared to White teachers, Black teachers in the sample also had higher teacher effects on students' math test scores in years prior to the experiment (0.01 versus -0.06 SD, for Black and White teachers, respectively), which is consistent with the experimental findings I show below.

To inform generalizability, in Table 1, I further compare teachers and students in the experiment to broader populations. The subset of volunteer teachers in the experiment look similar to the full NCTE project sample (n = 321) on a range of background teacher characteristics (i.e., gender, race/ethnicity, teaching experience), with no statistically significant differences. I also show that the subset of teachers in the experimental sample look similar to all upper-elementary teachers in their respective districts on their impacts on students' math test scores (p = 0.687). Background teacher characteristics were not available for all teachers in the four partner districts. I observe statistically significant differences across samples on several student characteristics. However, the magnitudes of these differences tend to be small. Roughly 40% of students were Black, 25% were Hispanic, 20% were White, and 10% were Asian.

Data

The analyses presented in this paper draw on a combination of primary data collected as part of the research project and secondary data collected from districts' administrative records that, together, illustrate varied social interactions and outcomes within schools and classrooms. I organize discussion around student- and teacher-level measures.

Students' Social-Emotional, Academic, and Behavioral Outcomes

I examine seven total student outcomes, which I categorize into three buckets related to intrapersonal competencies, test-based achievement, and observed school behaviors. At the end of the school year, research project staff administered a survey capturing three intrapersonal components of students' social-emotional development. These constructs align with theory on students' sense of self and place in the classroom that are thought to increase when assigned to a same-race/ethnicity teacher. *Self-Efficacy* (10 items, internal consistency reliability [α] = 0.76) captures students' effort, initiative, and perception that they can complete tasks. *Engagement and Happiness in Class* (5 items, α = 0.82) asks students about their affect, happiness in, and enjoyment of class activities.[3] And, *Self-Regulation* (3 items, α = 0.74) captures the extent to which students regulate their behavior to align with teachers' expectations. (See Table 2 for descriptive statistics, Appendix Table 1 for item text, and Blazar & Kraft, 2017 for exploratory factor analyses.) For each of these student-reported outcomes, I created final scales by reverse coding items with negative valence, averaging student responses across all available items within the construct, and then standardizing to mean of 0 and SD of 1.

District administrative data include student demographic characteristics, scores on state math and ELA tests federally mandated for annual reporting, and observed school behaviors (e.g., absences,

[3] *Self-Efficacy* and *Engagement and Happiness in Class* are strongly correlated (ρ = 0.65; see Table 2). However, review of item texts against the psychology literature from which they were derived provides theoretical justification for dividing them into two distinct constructs (see Blazar & Kraft, 2017). Sufficiently high reliability for each construct also supports this decision. Finally, estimated effects of Black versus White teachers and mediating pathways on these outcomes are similar, but tell slightly different stories.

suspensions). These data were available during the research study when students were in upper-elementary school (i.e., fourth or fifth grade during the 2012-13 experiment) and all subsequent years through 2018-19, allowing me to track outcomes up to six years later when students were in high school. I standardized all test scores within district, grade, and year to have a mean of 0 and a SD of 1. Absences data tally the total number of days students were absent from school each year, and suspension data capture the number of days that students were suspended from school (either in-school suspension or out-of-school suspension). On average, upper-elementary students in the experiment missed 6.2 school days and were suspended a total of 0.09 days; in high school, students missed 10.2 days and were suspended 1 day, on average. However, most students never were suspended, and many did not miss any days of school. Given the highly skewed nature of the absence and suspension data, I follow others in creating dichotomous measures that capture chronic absenteeism (i.e., missing 10% of total school days) and whether or not students were suspended at all in a given year (Gottfried, 2014; Holt & Gershenson, 2019; Lindsay & Hart, 2017; Jackson, 2018). For longer-run test scores, absences, and suspensions, 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. In upper-elementary school, 5% of students in the experimental sample were chronically absent and 4% were ever suspended; in high school, 17% of students were chronically absent and 18% were ever suspended (see Table 2).

In Table 2, I show that the student-reported constructs capturing intrapersonal competencies related to social-emotional development correlate with test-score performance, absences, and suspensions in expected directions, providing evidence of face validity. All three survey measures positively predict math and ELA test scores, with the highest correlation between *Self-Efficacy* and math test scores (ρ = 0.31). The student-reported measures also negatively correlate with absences and suspensions. *Self-Regulation* correlates most strongly with suspensions (ρ = -0.18), which makes sense

given that both aim to capture students' school and classroom (mis)behavior. Similarly, *Engagement and Happiness in Class* correlates more strongly with absences than with suspensions, which aligns with discussion of absenteeism as a proxy measure for students' engagement in school (Gottfried, 2014). In Appendix Table 2, I show that correlations between outcomes captured in upper-elementary versus high school are similar to those discussed above. For example, *Self-Efficacy, Engagement and Happiness in Class,* and *Self-Regulation* from elementary school all positively correlate with high school test scores (ρ = 0.15 to 0.27). *Self-Regulation* negatively correlates with chronic absenteeism (ρ = -0.11) and ever suspend (ρ = -0.16). These patterns align with similar evidence showing the predictive validity of student-reported social-emotional development to longer-run outcomes in college and the labor market (Lyubomirsky et al., 2005; Mueller & Plug, 2006).

Mindsets and Practices Aligned to "Good" Teaching

Data on teachers come from a survey administered each fall during the research project, as well as videotaped observations of classes collected throughout the school year. For this paper, I select a subset of constructs from the larger NCTE research project and dataset because of their alignment to theoretical discussion on "good" teaching practices and mindsets likely to show up in the classrooms of Black teachers. Because the research project was not designed specifically to capture culturally relevant or culturally responsive teaching, components related to "opposition" pedagogies (e.g., critical consciousness) are not available.

From the teacher survey, I focus on three teacher-reported mindset and practice measures. First is teachers' *Growth Mindset Beliefs* (7 items, α = 0.82), which captures the extent to which teachers view student intelligence as malleable versus fixed (Dweck, 2006). The former perspective is thought to be most beneficial for students because it can support students' own development of a growth mindset (Mesier et al., 2021) that, in turn, supports longer-run outcomes and academic success (Yeager & Dweck, 2020). Thus, I construct the *Growth Mindset Beliefs* measure such that higher scores reflect a

mindset that student intelligence is more malleable. While this construct stems most directly from the psychology literature, there also are direct connections with the culturally relevant pedagogy literature, which describes how teachers should hold and then act on beliefs that "knowledge is not static" and that all students are capable of academic success no matter their prior background, experiences, and circumstances (Ladson-Billings, 1995b, p. 481). (See Table 3 for descriptive statistics, and Appendix Table 3 for survey item text.)

Second is teachers' *Preparation for Instruction* (14 items, $\alpha = 0.78$), which identifies the amount of time teachers spend planning for instruction and collecting formative assessment data, as well as the extent to which they use this information to deliver differentiated instruction that attends to individual students' needs. Scaffolding instruction is described in the early literature on culturally relevant pedagogy (Ladson-Billings, 1995a). More recently, scholars discuss how approaches to differentiation and data-driven instruction have developed in tandem with culturally relevant pedagogies, all of which aim to better support historically marginalized and minoritized students to achieve academic success (Kieran & Anderson, 2019; Santamaria, 2019). The third construct from the teacher survey is *Relationships with Students and Families* (4 items, $\alpha = 0.63$), including 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. For example, one item asks teachers the extent to which "students and I show an interest in each other's lives." This and the other three items from this measures align closely with discussion of culturally relevant and culturally responsive teaching, grounded in strong social relations and instruction developed around students' lived experiences and backgrounds. For all teacher-reported constructs, I averaged response across all available years of data, and standardized measures to have a mean of 0 and standard deviation of 1.[4]

[4] In order to align the teacher survey constructs to the theoretical literature, I make several decisions and small changes to the originally designed items and constructs. First, for *Growth Mindset Beliefs*, three of the seven items were collected in one school year and four were collected in a different school year. Items were updated over time in light of initial psychometric

In addition to completing surveys, teachers contributed an average of three videotaped lessons per school year[5], which trained raters scored on the Classroom Assessment Scoring System (CLASS) instrument (Pianta et al., 2012). Exploratory and confirmatory factor analyses of CLASS scores in the data used here identify two constructs (Blazar et al., 2017) that both align with discussion of "good" teaching practices in the theoretical literature on teacher-student race/ethnicity matching. *Classroom Support* (9 items, $\alpha = 0.90$) 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.[6] *Classroom Organization* (3 items, $\alpha = 0.72$) captures teachers' behavior management skills and the extent to which teachers' approach to addressing student (mis)behaviors avoids creating a negative classroom culture. Teachers' ability to productively organize and respond to student (mis)behavior is described as particularly important to building culturally responsive classrooms given histories of exclusionary discipline for students of

analyses ($\alpha = 0.56$ for the first set of items, and 0.93 for the second set of items). I calculate internal consistency reliability across all items by considering responses for teachers who completed the survey in both school years, and collapsing data to the teacher level. I include all seven items given the high pooled reliability estimate, and because use of all items leads to higher coverage across teachers in the sample. Patterns of results are similar when excluding items from one year with lower reliability.

Second, the *Preparation for Instruction* items were developed to capture two constructs: out-of-class preparation and formative assessment. As shown in Appendix Table 3, each had its own response scale. 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). Finally, 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 (see Appendix Table 3). 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] Given the focus of the research study on mathematics teaching, all observations of classes come from math lessons. At the same time, all participating teachers were generalists who taught all core subjects. Capture occurred with a freestanding, three-camera, digital recording device and lasted roughly 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. Analyses from separate data indicate that teachers are ranked almost identically when they choose lessons versus when lessons are chosen for them (Ho & Kane, 2013).

[6] *Classroom Support* combines items from the *Emotional Support* (4 items) and *Instructional Support* (5 items) dimensions originally outlined by CLASS instrument developers. Confirmatory factor analyses of the data used in this study indicate that these items cluster together to form a single construct (Blazar et al., 2017).

color (Fenning & Rose, 2007), as well as misunderstandings of the behavior, physical movements, and language of minoritized students (Gay, 2002). (See Appendix Table 4 for item text.)

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 account for variation in the number of lessons teachers contributed to the dataset, I calculated predicted, shrunken teacher-level scores.[7] Using the raw data, I also calculated adjusted intraclass correlations (ICCs) that capture the amount of construct-relevant variation at the teacher level, as opposed to construct-irrelevant sources of variation including raters and time of year. Scores of 0.63 and 0.47 for *Classroom Organization* and *Classroom Support*, respectively, are very similar to other large-scale studies that use the CLASS and other observation instruments to score classroom instruction (Bell et al., 2012; Kane & Staiger, 2012) and, thus, provide evidence of score reliability in addition to internal consistency reliability. I standardized teacher-level observation scores to have a mean of 0 and a SD of 1.

In Table 3, I show that Black teachers outperformed White teachers on all five mindset and practice measures, with differences as large as 0.3 SD (i.e., *Growth Mindset Beliefs*). These differences are estimated from regression models that include school and grade fixed effects in order to compare teachers to others in the same randomized group (see Empirical Strategy section below for details). Between-group differences align with theory and motivate analyses that examine whether these mindset and practice measures mediate effects of Black versus White teachers on student outcomes.

[7] To estimate shrunken teacher observation scores, I specified the following multilevel model:

$$Observation_{lj} = \gamma_j + \varepsilon_{lj}$$

The outcome is the score for lesson *l* from teacher *j*. Teacher random effects, γ_j , are the parameters of interest, capturing mean observation scores for each teacher, shrunk to the sample mean based on the number of lessons for that teacher. From this model, I also calculate intra-class correlations (ICC) that estimate the amount of variance in scores attributable to the teacher. Following a generalizability study framework (Hill et al., 2012), I adjust the ICCs for the median number of lessons per teacher.

Empirical Strategy

In order to examine how and why Black teachers benefit the outcomes of their students, I draw on a causal mediation framework that consists of three equations (Baron & Kenny, 1986; Imai et al., 2010; VanderWeele, 2015):

$$Y^{c}_{isgj(t+n)} = \beta_1 BlackTch_{ijt=0} + \beta_2 OtherRaceTch_{ijt=0} + \beta_3 \boldsymbol{x}_i + \omega_s + \gamma_g + \varepsilon_{isgjt} \quad (1)$$

$$M^{m}_{ij(t=0)} = \delta_1 BlackTch_{ij(t=0)} + \delta_2 OtherRaceTch_{ijt=0} + \delta_3 \boldsymbol{x}_i + \omega_s + \gamma_g + \epsilon_{isgjt} \quad (2)$$

$$Y^{c}_{isgj(t+n)} = \pi_1 BlackTch_{ij(t=0)} + \pi_2 OtherRaceTch_{ijt=0} + \pi_3 \boldsymbol{x}_i + \pi_3 M^{m}_{ij(t=0)} + \omega_s + \gamma_g + \mu_{isgjt} \quad (3)$$

Equation (1) estimates the *total effect* of Black versus White teachers on outcomes, $Y^{c}_{isgj(t+n)}$, for student *i* in school *s* and grade *g* randomly assigned to teacher *j*, with each social-emotional, academic, or behavioral measure denoted by the superscript *c*. The time subscripts $t + n$ indicate that outcomes were 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. The main independent variable, $BlackTch_{ij(t=0)}$, captures whether or not students' randomly assigned teacher was Black. To isolate comparisons between Black and White teachers—with White teachers as the left-out category—I control for $OtherRaceTch_{ijt=0}$, indicating whether or not a student's teacher was neither Black nor White. Here and in other equations, I control for fixed effects for school, ω_s, and grade, γ_g, to match the blocked randomized design.[8] In most models, I also condition on a set of background student characteristics, $\boldsymbol{x}_i$, in order to increase the precision of estimates. I calculate robust standard errors clustered at the teacher level to account for the clustered randomized design, with students nested within teachers' classrooms.

[8] In the experimental design, randomization blocks are equivalent to school-grade combinations. I expand identifying variation—by controlling for fixed effects for school and grade, rather than school-grade—given that not all school-grades with a Black teacher also had a White teacher; comparatively, all but one school with a Black teacher also had a White teacher. Patterns of results are similar and lead to similar conclusions when I control for school-grade fixed effects rather than school and grade fixed effects.

Equation (2) estimates differences in a given teacher mindset or practice measure, $M^{m}_{ij(t=0)}$, between Black and White teachers, with each hypothesized mediator denoted by the *m* superscript. Estimates from equation (2) can be thought of as students' exposure to a given mindset or practice measure when randomly assigned to a Black teacher versus a White teacher. In Table 3 described above, I show statistically significant and substantively meaningful differences between Black and White teachers on all five hypothesized mediators. I construct the mediator measures using data captured up to start of the random assignment year (i.e., $t = 0$) in order to ensure that the measures were not influenced by the set of students in the classroom post randomization, as has been observed in other studies (Steinberg & Garret, 2016). This strategy also meets causal mediation assumptions of temporal ordering (Valeri & VanderWeele, 2013), as students were exposed to teachers' race prior to their mindsets and practices.

Finally, equation (3) captures the *direct effect* of Black versus White teachers on student outcomes, parsed from the effect of the hypothesized mediators. Thus, the difference between the direct effect from equation (3), π_1, and the total effect, β_1, from equation (1) is sometimes interpreted as the *mediated* or *indirect effect* (VanderWeele, 2015). If this difference is zero, then there is no mediating pathway. This also implies that the mindset or practice measure is unrelated to student outcomes (or that there is no difference in the measure between Black and White teachers, which is not the case here; see Table 3). If the difference between π_1 and β_1 is large and negative, this is indication of mediation because the mindset or practice measure reduces the magnitude of the total effect and, thus, explains some of the effect of Black versus White teachers on the outcome. In the results presented below, I consider the percent of the total effect that is explained by the mediated effect, or π_1 minus β_1 divided by β_1.[9]

[9] The literature on causal mediation suggests that, in some instances, it can be useful to specify a fourth equation that includes an interaction between the hypothesized mediator and the treatment indicator (MacKinnon et al., 2020;

For equations (1) and (3) that predict student outcomes, I augment the models to interact teacher and student race/ethnicity in order to separate average total or direct effects from race-matching total or direct effects. The theoretical and empirical literatures indicate that teacher-student race-matching may be more potent than average effects of teachers of color across students (Redding, 2019), driven in part by role-modeling channels (Fordham & Ogbu, 1984; Villegas & Lucas, 2004). It may also be the case Black teachers benefit the outcomes of all students, particularly if Black teachers tend to engage in "good" teaching practices more often than White teachers. The interaction models allow me to estimate subgroup effects of Black versus White teachers separately for Black students and for non-Black students (i.e., White, Hispanic, Asian; see Table 1), and then to test for differential effects across these two student groups. To generate these estimates, I interact the three race/ethnicity categories for teachers (i.e., Black, White, non-Black/non-White) and all student background characteristics with two student race/ethnicity categories (i.e., Black, non-Black). I focus on just two student subgroups for both substantive and practical reasons. If the effects of Black teachers are driven by race-matching, then the important comparison is between Black students versus students from any other racial/ethnic background. Further, the sample size and statistical power decrease substantially when estimating effects of Black versus White teachers on narrower subgroups of students. That said,

VanderWeele, 2015). There are several technical and substantive reasons for this approach that do not apply to the current analyses. First, inclusion of an interaction allows for estimation of the *controlled direct effect,* which helps avoid bias that often emerges when conditioning on post-treatment variables (i.e., the mediators) that often are endogenous. This is, even in an experiment, the mediators generally are not randomly assigned and so can result in what Archaya et al. (2016) refer to as "intermediate variable bias". This concern can be avoided when focusing on the controlled direct effect because it represents the causal effect of a treatment when the mediator is fixed at a particular level. However, intermediate variable bias is unlikely to be a concern in the current context: when teachers are randomly assigned to classes, variation in pre-experimental teacher race/ethnicity *and* in mindset and practice mediators is exogenous. Thus, inclusion of an interaction is not necessary for causal identification. Second, inclusion of an interaction between treatment status and mediator(s) can be substantively meaningful in a policy context that aims to intervene on both the treatment and the mediator(s) (i.e., fix the treatment and the mediator to particular levels). However, that is not the policy goal in the current analyses. As I describe at the end of the paper, the presence versus lack of mediating pathways likely suggest different policy solutions: in the case of little or no mediation, the recruitment of Black teachers is the key policy solution because their effects on students are likely to be driven by role-modeling effects rather than through teaching skills; comparatively, in the case of substantial mediation, recruitment of Black teachers may be supplemented by professional development aimed at supporting all teachers to hold mindsets and engage in practices that drive effects on student outcomes. While both approaches may occur simultaneously, they require different strategies and target different populations.

in the Discussion section, I describe effects of Black versus White teachers on White students—in an exploratory way—in order to benchmark findings against other studies that make this comparison.

In order to interpret the set of effects described above causally, several assumptions need to be met aligned to what Imai et al. (2010) call "sequential ignorability" (for additional discussion of causal mediation assumptions, see Archaya et al., 2016; MacKinnon et al., 2020; VanderWeele, 2015). There are two ignorability assumptions, which are made sequentially: (i) treatment assignment is ignorable (i.e., statistically independent) of potential outcomes and potential mediators; and (ii) the mediators are ignorable, conditional on the treatment and pretreatment confounders. The first assumption is central to all causal analyses—mediation focused or otherwise—and is reasonable in the current analysis given the randomized design. In Table 4, I show that background characteristics of students are balanced across the classrooms of Black versus White teachers ($p = 0.338$ on joint test of significance). Estimates come from a regression model that predicts whether or not students' randomly assigned teacher was Black as a function of all available student characteristics, including demographics and prior test scores and school behaviors, as well as school and grad fixed effects; I also control for whether or not the teacher is neither Black nor White to isolate a comparison between Black and White teachers.[10]

The second part of the sequential ignorability assumption states that the mediators can be regarded is if they were randomized, conditional on treatment status. In general, this assumption is considered to be quite strong, even in the context of a randomized trial. For example, the grounding literature for this study suggests that access to a Black teacher may benefit students' academic performance through intermediate effects on students' social-emotional development (Ferguson, 2003; Gay, 2000). The randomized design allows me to estimate causal effects of Black versus White

[10] Treatment assignment is not statistically independent of the mediators, as the teacher mindset and practice measures are characteristics of and follow Black teachers. That said, this feature of the data helps meet the second assumption of sequential ignorability.

teachers on test scores and, separately, on intrapersonal components of students' social-emotional development, as shown in equation (1). But, it is more difficult to test how effects of Black versus White teachers on social-emotional development translate into effects on academic performance, since social-emotional development is not randomly assigned to students. For this reason, I do not estimate these pathways.

That said, this paper is primarily interested in teacher- rather than student-level mediators, and the random assignment of teachers to classes means that *all* background teacher characteristics should be statistically independent of background student characteristics. I provide empirical support for this assumption in Table 4 by specifying regression models that predict each teacher-level mindset or practice measure—captured at baseline, as noted above—as a function of student characteristics, school and grade fixed effects, and dummy indicators for teachers' race/ethnicity. While I find some marginally statistically significant relationships in some models, estimates from a multivariate multiple regression model that generates all estimates simultaneously returns no relationship between background student characteristics and the full set of teacher-level variables ($p = 0.390$).[11] Another way to test this assumption proposed by Imai et al. (2011) is to correlate residuals from equation (2)—the mediation model—and from equation (3)—the outcomes model that includes the mediators. Sequential ignorability holds overall when the correlations are zero, which I find is the case across all of the mediators.

An additional point about assumptions and causal estimation is the role of missing data. Even if student characteristics are balanced across teachers and classrooms at baseline, bias still could be introduced if there were differential attrition. In this study, student-level attrition was driven by

[11] Multivariate multiple regression analyses do not easily accommodate clustering of standard errors. Therefore, I conduct the analyses in a class/teacher-level dataset, with background student characteristics averaged to the class level. I also run analyses in a teacher-level dataset without clustering of standard errors, and find no relationship between the set of background student characteristics and the set of teacher-level variables ($p = 0.623$).

missing data due to three reasons. First, 32% of students moved out of their randomly assigned teachers' classroom, of whom 79% did not participate in primary data collection of student surveys.[12] Aligned to an intent-to-treat framework, I still include noncompliers in the analytic sample as long as they have outcome data. However, I cannot include individuals with missing outcome data. Second, some students were not visible or had missing data in school districts' records capturing test scores, absences, and suspensions. During the random assignment year, 11% of students were missing test scores and 21% were missing absences and suspensions.[13] In the longer-run high school data, 48% were missing test scores and 42% were missing absences and suspensions, driven largely by students moving out of their original district between upper-elementary school and high school. In addition, 14% of teachers were new to the study in the experimental year and so are missing values of all mediators that were captured at baseline.

In Appendix Table 5, I show that missingness is not systematically related to the treatment nor to background characteristics of individuals, and so should not introduce any threats to internal validity. At the student level, differentials in the proportion of individuals with outcome data between those randomly assigned to a Black versus a White teacher are small (1 to 6 percentage points), and none of these differences is statistically significant. Further, the background characteristics of students (see Table 1) with versus without outcome data do not differ for those individuals whose randomly assigned teacher was Black versus White. To examine these relationships, I specify a set of linear probability regression models that predict whether or not students' randomly assigned teacher was

[12] Of student non-compliers, 62% moved to another classroom within the same school, 17% switched schools within the same district, 18% moved out of the district, and 4% moved for unknown reasons. Importantly, though, these moves were unrelated to whether or not their assigned teacher was Black versus White ($p = 0.745$).

[13] During the experimental year, more students were missing absences and suspension records than test scores given the process of assembling district administrative records. Short-term test scores come from the original data files collected by the NCTE study, but these records do not include absences and suspensions. When districts provided the longer-run outcome data in high school, they also provided earlier data on absences and suspensions. However, some students in the original study could not be linked in subsequent data requests because they eventually dropped from the study and so do not show up in the crosswalk between study and district identification numbers.

Black as a function of interactions between having a specific data source and all background characteristics of the student, as well as their main effects; and controlling for a dummy indicator for whether or not students' randomly assigned teacher was non-Black/non-White (with White teachers as the left-out category), and school and grade fixed effects. *P*-values on joint tests of significance that the interaction terms are simultaneously equal to zero are above standard thresholds of statistical significance across outcome measures ($p > 0.18$). Thus, in all analyses, I exclude students with missing outcome data. At the teacher level, the differential between Black and White individuals in terms of missing mediator data also is small (6 percentage points) and not statistically significant; and, the background characteristics of teachers with missing mediator data are similar between Black and White teachers ($p = 0.357$ on joint test of significance). Rather than drop all observations of the students of these teachers, I impute values of the mediators to the school mean and, as noted above, include school fixed effects in all models.

Results

Total Effects of Black versus White Teachers

In Table 5, I present estimates of the total effect (without accounting for the mediators) of Black versus White teachers on the set of short- and longer-term student outcomes. In the first two columns, I estimate effects on average across all students, no matter their race or ethnicity. Estimates in each cell come from separate regression models that control for school and grade fixed effects, and a dummy indicator for whether or not students' randomly assigned teacher was neither Black nor White. Estimates in the second column add background student characteristics. Given the randomized design and baseline balance (see Table 4), estimates of the effect of Black versus White teachers on student outcomes should be very similar when including versus excluding the student-level controls. I find that this is the case. All subsequent models in this table and others include the student controls because they help increase precision (i.e., smaller standard errors). In the next two columns, I present

subgroup effects of Black versus White teachers on the outcomes of Black students and non-Black students. These estimates—and differences in their magnitude—come from a single regression model where teacher and student race/ethnicity categories are interacted. Chronically absent and ever suspended variables are binary; thus, estimates from linear probability models describe the percentage point increase in these outcome measures when having a Black versus a White teacher. All other outcomes are standardized.

I find very large effects of Black versus White teachers on intrapersonal components of social-emotional development that tend to be localized to Black students. Assignment to a Black versus a White teacher increases Black students' self-reported *Self-Efficacy* by 0.56 SD and their *Engagement and Happiness in Class* by 0.66 SD. The magnitude of the race-matching effect on *Self-Regulation* also is substantively meaningful (0.21 SD) but not statistically significantly different from zero. The effect of Black versus White teachers on the *Self-Efficacy* of non-Black students is similar in magnitude (0.18 SD) but also not statistically significantly different from zero. I can detect differences in the effects of Black versus White teachers on Black versus non-Black students on the *Self-Efficacy* and *Engagement* measures, while the between-group difference for *Self-Regulation* is close to but not lower than standard thresholds ($p = 0.119$). Standard errors for effects on the student-reported measures are larger than for effects on other outcomes, likely due to greater measurement error (see Table 2); measurement error in dependent variables does not lead to bias, but does decrease precision.

Black upper-elementary teachers also impact end-of-year test scores and school behaviors, and many of these effects extend to all students. Because the subgroup effects on these short-term outcomes are not statistically significantly different from each other, I focus on interpreting the average effects across all students. Assignment to a Black versus a White teacher improves math achievement by 0.12 SD and ELA achievement by 0.2 SD, and reduces chronic absenteeism by 4.4 percentage points. Compared to a chronic absenteeism rate for the students of White teachers of 6.9

percent (not shown in Table 5), the estimated effect on chronic absenteeism represents a roughly 65% reduction. For chronic absenteeism and ELA achievement, these short-term, end-of-year effects also persist at fairly similar magnitudes up to six years later when students were in high school. The longer-run effects on ELA achievement (0.27 SD) are concentrated amongst non-Black students, while the longer-run effects on chronic absenteeism (11.8 percentage points or 61% reduction relative to rate for students of White teachers) are concentrated amongst Black students. (Student-reported measures were not available in high school.)

Findings related to suspensions may appear counterintuitive: I find that assignment to a Black teacher results in an increased likelihood of being suspended during that year (roughly 6 percentage points), compared to assignment to a White teacher. Notably, though, I do not find differential suspension rates of Black versus non-Black students assigned to a Black teacher ($p = 0.612$). In contrast, I do find differential suspension rates of Black versus White students assigned to a White teacher (3.7 percentage points $p = 0.037$; not shown in Table 5), which is consistent with other analyses looking at exclusionary discipline for Black students.

Mediating Pathways Through Teachers' Mindsets and Practices

Next, in Table 6, I examine the extent to which the total effects of Black versus White teachers on the outcomes of their students are explained by the set of hypothesized mediators capturing "good" teaching mindsets and practices. In the main analyses, I focus on mediating effects on average across all students, rather than mediating pathways for student subgroups for three reasons. First, in the absence of differential effects between Black and non-Black students, it is appropriate to examine mediating pathways for average effects across all students. Second, when I do observe differential effects between Black versus non-Black students, main effects are concentrated in one group or the other. Thus, mediating pathways for the average effects also are driven by one group versus the other, rather than there being differential mediating pathways across student groups. Relatedly, as shown in

Appendix Table 7, in instances of subgroup differences in the total effects, mediating pathways for one student group or the other are very similar to mediating pathways for the average effects that I focus on in the main results.

The primary estimates of interest in Table 6 are percent changes in the total effect of Black versus White teachers from models that exclude mediators relative to models that account for one or multiple mediators. The percent change is the direct effect—estimated from equation (3)—minus the total effect—estimated from equation (1)—divided by the total effect. The total effects listed in this table are the same as those presented earlier and are included as a point of reference. In the first of these columns assessing mediation, I include all teacher mindset and practices in the same model. Then, I show estimates from models that add the hypothesized mediators one at a time. For simplicity, I do not show the direct effects because they can be inferred from the total effect and the percent change. Instead, I show estimates from the full mediation model—including the direct effects and the relationship between mediators and student outcomes—in Appendix Table 6.

Given the large number of percent change estimates presented, I use shading and bolding to guide the reader. I shade rows in gray in instances where the total effect of Black versus White teachers is statistically significant, as mediation only can occur when there is a total effect. However, I still show estimates for other outcomes where the total effect is substantively meaningful, even if not statistically significantly different from zero (i.e., short-term *Self-Regulation* and longer-term chronic absenteeism). I do not show change estimates for longer-run math achievement and suspensions, where the total effects are very close to zero and so have very little meaningful interpretation. The change estimates also explode because they are derived by dividing the difference between the total and direct effects by the total effect that is close to zero. Then, I bold percent change estimates that are likely to be substantively meaningful based on magnitude (5% change or greater) and sign. In general, a negative percent change value is more desirable, as it indicates that the hypothesized mediator is observed more

for Black teachers versus White teachers (see Table 3) and that the mediator is associated with improved student outcomes (see Appendix Table 6). An exception, though, is for the effect of Black versus White teachers on short-term suspensions, where a positive percent change is more desirable given that the total effect goes in the "wrong" direction (i.e., increased suspensions). Here, a positive percent change indicates a positive relationship between the mediator(s) and the outcome, and Black teachers outperforming White teachers on the mediator(s). Thus, I bold percent change estimates for short-term suspensions that are positive and greater than five.

There are two useful ways to interpret causal mediation in Table 6: looking across the rows to assess mediation on a given outcome, and looking down the columns to assess mediation from an individual mediator. Starting with the latter approach, I find that *Classroom Organization* has the largest mediating effect. This classroom practice accounts for 23% of the total effect of Black versus White teachers on students' short-term math achievement. I observe a larger percent change on student-reported *Self-Regulation* (49%), which makes sense given the close alignment between these teacher and student measures that both focus on behavior and organization in the classroom. While this mediating pathway is the strongest across all results presented in this paper, it should be interpreted with caution given that the total effect of Black versus White teachers on students' *Self-Regulation* is not statistically significantly different from zero. Teachers' *Growth Mindset Beliefs* that student intelligence is malleable rather than fixed also mediates the total effect of Black versus White teachers on several outcomes. The percent change is smaller than for *Classroom Organization*, but still meaningful: 7% for student-reported *Engagement*, 9% for short-term suspensions, and 12% for chronic absenteeism in high school. The patterns indicate that *Growth Mindset Beliefs* largely explain effects of Black versus White teachers on engagement-related student outcomes.

Looking next across the rows further shows that individual mediators—rather than the set of mediators as a group—generally drive changes in the total effect of Black versus White teachers on a

given outcome. The largest degree of mediation occurs for *Self-Regulation* (56%) and short-term math achievement (16%). In both instances, there is one key driver: *Classroom Organization*. For effects on short-term on *Engagement* and longer-run chronic absenteeism, the total amount of mediation is driven primarily by *Growth Mindset Beliefs*, though most other mediators play a smaller role. Notably, even though the total effect of Black versus White teachers is largest for student-reported *Self-Efficacy*, I do not find any meaningful mediation here using the available teacher mindset and practice measures.

An additional finding from Table 6 is the lack of mediating pathways for three additional mindset and practice measures. *Preparation for Instruction* only reduces the total effect of Black versus White teachers on student outcomes across all students to a very small degree (no more than 3%). *Classroom Support*, which is one of the two remaining mediators focused on teacher-student relationships, reduces the total effect of Black versus White teachers on short-term math achievement and longer-run chronic absenteeism to a modest degree (both 4%). But, this practice also is undesirable in other instances (e.g., associated with worse short- and longer-term ELA achievement and worse *Self-Regulation*; see Appendix Table 6). I observe similar contradictory patterns for a related measure, *Relationships with Students and Families*, which comes from a teacher survey as opposed to observations of classroom instruction. This measure is associated with worse short-term test scores in both subjects, but improved attendance at school. In analyses of mediating pathways for subgroup effects, mediating pathways through *Classroom Support* and *Preparation for Instruction* to longer-run chronic absenteeism of Black students are somewhat stronger (5% to 6%; see Appendix Table 7). In other words, while Black teacher outperform White teachers on these measures—these are not the pathways that drive improved student outcomes.

Discussion

Pairing random assignment of teachers to classes with rich data on varied student outcomes and varied teacher mindsets and practices, this study extends prior quantitative literature identifying

positive effects of teacher-student race/ethnicity matching, as well as primarily qualitative and theoretical literature characterizing the "pedagogical excellence" of Black teachers into a set of "good" teaching mindsets and practices thought to drive student outcomes.

Benchmarking Total Effects of Black versus White Teachers

A key finding from this study is that Black teachers have very large total effects on an array of student outcomes, many of which extend to all students (in addition to Black students), and persist at similar magnitudes several years after working with this Black teacher. The effects of Black versus White teachers on intrapersonal components of social-emotional development for their Black students are particularly strong (0.56 SD on *Self-Efficacy* and 0.66 SD on *Engagement*), and compare quite favorably to other analyses. In a highly cited meta-analysis examining impacts of school-based social-emotional development interventions, Durlak et al. (2011) found average effects on students' attitudes towards self and others (e.g., self-efficacy) of 0.23 SD and on positive social behaviors 0.24 SD; these outcomes are most similar to the ones assessed in the current study. Effects on a broader set of social-emotional skills (e.g., goal setting, perspective taking, problem solving) were larger at 0.57 SD. A more recent review of similar school-based interventions that narrows in on randomized control trials found average effects of 0.35 SD on social-emotional development and 0.28 SD on behavioral adjustment (Korpershoek et al., 2016). Estimates from the current analyses often are larger, suggesting that access to same-race teachers is a particularly important strategy for improving the self-efficacy and classroom engagement of Black students, just as the theoretical literature describes (Ferguson, 2003; Fordham & Ogbu, 1986; Gay, 2000; Villegas & Lucas, 2004). While effects on the student-reported measures are localized to Black students, I also find large decreases in chronic absenteeism of Black versus White teachers that extend to all students (roughly 65% reductions), likely capturing, in part, students' engagement in school (Gottfried, 2014).

I also identify effects of Black versus White teachers on student test scores, which are not the

main focus of the current analyses—primarily interested in a set of classroom social interactions—but are notable for several reasons. First, I find very similar end-of-year test-score effects (0.12 SD in math and 0.2 SD in ELA) as Dee (2004) and replicated by Gershenson et al. (2021), who draw on data from the Project STAR experiment in which teachers also were randomly assigned to classes. In comparison, a growing number of non-experimental studies that often exploit within-student changes in teacher race and outcomes over time find positive test-score effects of teacher-student race/ethnicity-matching, but at magnitudes that generally are much smaller (for reviews, see Bristol & Martin-Fernandez, 2019; Redding, 2019). For example, Egalite et al. (2015) used statewide data from Florida and find test-score effects of 0.02 SD for Black students matched with a Black teacher. Non-experimental analyses that draw on large statewide datasets have the benefit of increased generalizability, as compared to experiments with volunteer samples. However, these analyses likely suffer from internal validity threats due to the non-random sorting of students to teachers. Because teachers of color tend to be assigned to classes where students have below-average test-score performance and above-average absence and suspension rates (Holt & Gershenson, 2019; Kalogrides et al., 2013), non-experimental analyses likely understate true effects of teachers of color on student outcomes.

Preferencing experimental estimates from this study and the Project STAR data, the literature is clear that assignment to a Black versus White teacher produces some of the largest effects on the test-score outcomes of their students all of the education intervention literature. As a general point of comparison, Fryer (2017) conducted a meta-analysis of experimental effects of a broad range of human-capital oriented interventions on student test scores, finding the largest effects for one-on-one, high-dosage tutoring (0.31 SD in math and 0.23 SD in ELA). The effects of Black teachers on ELA performance are similar in magnitude, and they extend to a range of students in the classroom rather than being a one-on-one intervention like tutoring.

Second, unlike findings from the Project STAR experiment, I find that Black teachers improve the test-score performance—and some other outcomes—of non-Black students in addition to Black students. In other words, these are not just race-matching effects, but average effects of Black teachers. Specifically, I find that the impacts of Black versus White teachers on end-of-year math and ELA performance are quite similar and statistically indistinguishable across Black and non-Black students. (This pattern also is true for effects on short-term chronic absenteeism.) One potential reason for differences in findings between this study and others may be related to the set of non-Black students included in the sample. In mid 1980s Tennessee, the Project STAR sample consisted almost exclusively of Black and White students, leading Dee (2004) and Gershenson et al. (2021) to estimate effects of Black versus White teachers for these two student subgroups only. They found positive race-matching effects for Black students paired with a Black teacher, compared to assignment to a White teacher; and null or negative effects for White students paired with a Black teacher, compared to assignment to a White teacher. In comparison, the sample in this study includes greater variation in terms of student race/ethnicity—41% Black, 20% White, 24% Hispanic, and 10% Asian—and I constructed the subgroup of non-Black students to include the latter three. Thus, it may be that effects of Black versus White teachers on non-Black students are driven by non-Black students of color who also uniquely benefit from having a teacher of color. That said, when I set up analyses to look more similar to Gershenson et al. (2021), I find evidence that White students benefit from having a Black versus a White teacher in terms of decreased chronic absenteeism in upper-elementary school (7 percentage points; $p = 0.046$) and increased ELA test scores in high school (0.26 SD; $p = 0.029$). These patterns provide new evidence that Black teachers benefit all students, including White students.

A final notable pattern related both to test scores and to behavioral outcomes is their persistence over time. I find effects of Black versus White teachers on non-Black students' ELA achievement in high school of 0.27 SD, six years after working with this teacher. This pattern is almost

unheard of in education research, where the effects of teachers and school-based interventions on test scores—and on other measures—tend to fade out substantially over time (Bailey et al., 2017; Cascio & Staiger, 2012). I also find persistent reductions in chronic absenteeism of Black students between upper-elementary school (4.4 percentage points and 64% reduction relative to comparison group) and high school (12 percentage points and 61% reduction).

One explanation for persistence—particularly for the outcomes of Black students—is that assignment to a Black versus a White teacher meets what Bailey et al. (2017) refer to as the "trifecta" condition of skill building. Black teachers benefit student outcomes that are (i) fundamental, (ii) malleable, and (iii) unlikely to develop in the absence of the intervention. More specifically, the authors describe a set of trifecta skills that include "advanced" skills in mathematics and literacy, "academic motivation", and "self-concept for adolescents facing stereotype threat" (p. 18), which are exactly the short-term skills that improve for Black students when assigned to a Black teacher. Further, I show that the absence of the treatment and the counterfactual condition—where Black students were assigned to a White teacher—can be harmful. In these classes, Black students were suspended and experienced exclusionary discipline more than their White peers. In turn, Black students assigned to a White teacher may not easily catch up to Black students assigned to a Black teacher, particularly in terms of their self-image and academic motivation.

For persistence for the test-scores of non-Black students, the "trifecta" condition may also hold. While I do not observe effects of Black versus White teachers on the self-reported survey measures of non-Black teachers, I do find effects on test-scores and chronic absenteeism that are both quite large. It may be that the combination of effects across different types of outcomes translates over time into sustained effects on test-score performance. In support of this claim, it is noteworthy that the sustained test-score effects show up only in ELA and not math, where the skills tested in high school ELA assessments (e.g., reading comprehension, narrative writing) are thought to benefit from

multi-component interventions that address both academic and social skills. For example, in an evaluation of a curriculum targeting both social-emotional development and academic performance, Brackett et al. (2012) found larger effects on ELA test scores than math. In another experiment of a related intervention, McCormick et al. (2021) found that effects were sustained longer in ELA than in math. For non-Black students, work with a Black teacher in early adolescence may also serve as a "foot-in-the-door" (Bailey et al., 2017) by exposing them to a diverse set of adults and perspectives at a time of developmental transition. I cannot test this hypothesis, and so it is a topic for a separate experiment where students are randomly assigned to Black versus White teachers at different points in their schooling trajectories.

Implications of Mediating Pathways for Policy and Practice

How then can schools and districts, policymakers and practitioners act on these findings? There are at least two, not mutually exclusive directions. Both of these approaches seek to address the fact that the (re)assignment of students to Black teachers is not a policy or practice solution in and of itself, given that Black teachers are wildly underrepresented in the teacher workforce (U.S. Department of Education, 2019).

First, analyses of mediating pathways help identify areas of "good" teaching mindsets and practices that explain the effects of Black versus White teachers on student outcomes, and so are key focal areas for teacher training and development. Teachers' ability to productively organize the classroom environment in a way that is conducive to learning and student engagement explains up to 50% of the total effects of Black versus White teachers on test-score and student-reported behavior; and, teachers' growth mindset beliefs that student intelligence is malleable rather than fixed explains up to 12% of the total effects on engagement, absenteeism, and suspensions. Thus, there may be some or even substantial benefit of allocating professional development dollars towards shifting the mindsets of White teachers towards a growth-mindset perspective and to help them build productive

classroom environments. Experiments show that training interventions can improve these mindsets and skills across all teachers, further resulting in increased student test-score performance and classroom engagement, as well as decreased racial disparities in exclusionary discipline (e.g., Early et al., 2017; Gregory et al., 2016; Yeager et al., 2019). This suggestion also aligns with recent evidence on the positive effects of an ethnic studies curriculum aligned to culturally responsive teaching and that was implemented by teachers of all races and ethnicities in California (Bonilla et al., 2021; Dee & Penner, 2017). Building towards practice-based teacher training and development in these areas may help address concerns that "so little of [good teaching] seems to be occurring in the classrooms populated by African American students" (Ladson-Billings, 1995a, p. 159), and classrooms across the U.S. more generally (Kane & Staiger, 2012).

Second, teacher training and development needs to be paired with widescale recruitment of Black teachers in order to provide Black students with more role models who look like them. Identifying role modeling effects generally is quite challenging in causal, quantitative work, and largely requires ruling out other possible channels (Gershenson et al., 2021). Thus, in the current analyses, I cannot prove the existence of role modeling directly. However, the findings point in this direction. For no student outcome do the set of available mediators explain more than 50% of the total effect of Black versus White teachers, meaning that there exist other mediating pathways. Some of these unexplained pathways likely run through teacher mindsets and practices not available in the current dataset, particularly those related to opposition pedagogies discussed in the culturally relevant, culturally responsive, and culturally sustaining teaching literatures. This is an area of exploration for additional research. Role modeling also is a likely avenue driving the total effects of Black versus White teachers on student-reported self-efficacy—and potentially for other outcomes—where I observe no substantively meaningful mediating pathways. Social cognitive theory directly links role modeling, self-image, and self-efficacy (Bandura, 1986), and these relationships are thought to be particularly

powerful for minoritized populations (Ferguson, 2003; Villegas & Lucas, 2004). For non-Black students, I also find longer-run impacts on test scores with no mediating pathways. Thus, Black teachers likely serve an important social function for all students, above and beyond the mindsets and practices they bring into the classroom. In order to increase access to Black teachers, the only real solution is to recruit more of them into schools.

Conclusion

Knowing what the policy and practice goals are does not make them simple or straightforward to achieve. On the recruitment side, shifting demographics of the teacher workforce is a numbers problem at a bare minimum. Currently, roughly 80% of U.S. teachers are White, 7% Black, and 9% Hispanic, while the student population is roughly 45% White, 15% Black, and 25% Hispanic. 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, 2017) and current Black teachers often are underappreciated for their work (Carver-Thomas & Darling-Hammond, 2017; Fairchild et al., 2012; Griffin & Tackie, 2017). A promising avenue for addressing may be grassroots and community-led "grow-your-own" programs that look locally for prospective teaching talent (e.g., Bristol & Goings, 2019; Dixon et al., 2019). But, at least at present and on the relatively small scale they are being implemented, these programs have not been associated with the magnitude of demographic change needed in the teacher workforce.

Further, while I advocate for training of the current teacher workforce that is mostly White, implementing successful programs at scale poses other challenges. Professional development programming such as one-on-one coaching can improve dimensions of "good" teaching that drive student outcomes in this study. Yet, their effects generally decline as programs grow in size (Kraft et al., 2018). Further, in light of racial biases instantiated in differential suspension rates of the students

of White teachers documented here and elsewhere (e.g., Fenning & Rose, 2007; Shirrell et al., 2021), I cannot rule out the possibility that some of 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. The fact that exclusionary discipline decreases amongst teachers who take part in researcher-led training (e.g., Gregory et al., 2016) does not mean that patterns will extend to broader populations. In fact, those teachers most in need of change may be the least likely to opt into the program.

Another challenge in acting on the findings from this paper is growing political tensions between progressives and conservatives around the role of race in school. Right-wing extremist arguments grounded in White supremacy suggest that White students will feel uncomfortable when exposed to any topic adjacent to critical race theory—including social-emotional learning—that even raises the notion of race-based inequities. These suggestions are being mainstreamed into school board elections, parent stakeholder meetings, and choices of books and curricula (Sawchuk, 2021). Thus, widespread efforts to shift teacher demographics and to train (mostly White) teachers in "good" teaching practices aligned to cultural responsiveness are certain to and already have received pushback. As a progressive (and beyond my role as a researcher), I advocate wholeheartedly for centering race, ethnicity, and culture in schools, and to advance policies specifically designed to better serve Black, Hispanic, and other students of color. It also is true that assignment to a Black teacher improves the outcomes of *all* students, including White students. This scientific fact will not change the minds of extremists, but may be a winning argument for those on the margin.

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 across all of the education research literature. These effects show up in test scores, social-emotional development, and school behaviors; they are persistent over time; and they extend to students no matter their racial or ethnic background. We must use these findings to compel large changes in policy and in practice.

References

Acharya, A., Blackwell, M., & Sen, M. (2016). Explaining causal findings without bias: Detecting and assessing direct effects. *American Political Science Review*, *110*(3), 512-529.

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.

Bandura, A. (1986). *Social foundations of thought and action*. Englewood Cliffs, NJ: Prentice Hall.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173-1182.

Bell, C. A., Gitomer, D. H., McCaffrey, D. F., Hamre, B. K., Pianta, R. C., & Qi, Y. (2012). An argument approach to observation protocol validity. *Educational Assessment*, *17*(2-3), 62-87.

Blazar, D., Braslow, D., Charalambous, C.Y. and 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.

Bonilla, S., Dee, T. S., & Penner, E. K. (2021). Ethnic studies increases longer-run academic engagement and attainment. *Proceedings of the National Academy of Sciences*, *118*(37).

Bottiani, J. H., Larson, K. E., Debnam, K. J., Bischoff, C. M., & Bradshaw, C. P. (2018). Promoting educators' use of culturally responsive practices: A systematic review of inservice interventions. *Journal of Teacher Education, 69*(4), 367-385.

Brackett, M. A., Rivers, S. E., Reyes, M. R., & Salovey, P. (2012). Enhancing academic performance and social and emotional competence with the RULER feeling words curriculum. *Learning and Individual Differences*, *22*(2), 218-224.

Bristol, T. J., & Goings, R. B. (2019). Exploring the boundary-heightening experiences of Black male teachers: Lessons for teacher education programs. *Journal of Teacher Education*, *70*(1), 51-64.

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.

Carver-Thomas, D., & Darling-Hammond, L. (2017). Why Black women teachers leave and what can be done about it. In A. Farinde-Wu, A. Allend-Handy, & C.W. Lewis (Eds.), *Black female teachers: Diversifying the United States' teacher workforce* (pp. 159-183). Emerald Publishing Limited.

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.

Chin, M. J., Quinn, D. M., Dhaliwal, T. K., & Lovison, V. S. (2020). Bias in the air: A nationwide exploration of teachers' implicit racial attitudes, aggregate bias, and student outcomes. *Educational Researcher*, *49*(8), 566-578.

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.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter?. *American Economic Review*, *95*(2), 158-165.

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

Dee, T., & Penner, E. (2019). *My brother's keeper? The impact of targeted educational supports* (No. w26386). National Bureau of Economic Research.

Delhommer, S. (2022). High school role models and minority college achievement. *Economics of Education Review*, *87*.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, *132*(4), 1593-1640.

Dixon, D., Griffin, A., & Teoh, M. (2019). *If you listen, we will stay: Why teachers of color leave and how to disrupt teacher turnover.* Washington, DC: The Education Trust and Teach Plus.

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.

Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random House Digital, Inc.

Early, D. M., Maxwell, K. L., Ponder, B. D., & Pan, Y. (2017). Improving teacher-child interactions: A randomized controlled trial of Making the Most of Classroom Interactions and My Teaching Partner professional development models. *Early Childhood Research Quarterly*, *38*, 57-70.

Eddy, C. M., & Easton-Brooks, D. (2011). Ethnic matching, school placement, and mathematics achievement of African American students from kindergarten through fifth grade. *Urban Education*, *46*(6), 1280-1299.

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.

Egalite, A. J., Kisida, B., & Winters, M. A. (2015). Representation in the classroom: The effect of own-race teachers on student achievement. *Economics of Education Review*, *45*, 44-52.

Fairchild, S., Tobias, R., Corcoran, S., Djukic, M., Kovner, C., & Noguera, P. (2012). White and black teachers' job satisfaction: Does relational demography matter?. *Urban Education*, *47*(1), 170-197.

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

Ferguson, R. F. (2003). Teachers' perceptions and expectations and the Black-White test score gap. *Urban Education*, *38*(4), 460-507.

Fordham, S., & Ogbu, J. U. (1986). Black students' school success: Coping with the "burden of 'acting white"'. *The Urban Review*, *18*(3), 176-206.

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. (2002). Preparing for culturally responsive teaching. *Journal of Teacher Education, 53*(2), 106-116.

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

Gershenson, S., Hart, C. M., Hyman, J., Lindsay, C., & Papageorge, N. W. (2021). *The long-run impacts of same-race teachers* (No. w25254). National Bureau of Economic Research.

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

Girvan, E. J., Gion, C., McIntosh, K., & Smolkowski, K. (2017). The relative contribution of subjective office referrals to racial disproportionality in school discipline. *School Psychology Quarterly*, *32*(3), 392.

Glazerman, S., Mayer, D., & Decker, P. (2006). Alternative routes to teaching: The impacts of Teach for America on student achievement and other outcomes. *Journal of Policy Analysis and Management*, *25*(1), 75-96.

Goffney, I. D. (2010). *Identifying, measuring, and defining equitable mathematics instruction* (Doctoral dissertation).

Gottfried, M., Kirksey, J. J., & Fletcher, T. L. (2021). Do High School Students With a Same-Race Teacher Attend Class More Often?. *Educational Evaluation and Policy Analysis*, 01623737211032241.

Gottfried, M. A. (2014). Chronic absenteeism and its effects on students' academic and socioemotional outcomes. *Journal of Education for Students Placed at Risk (JESPAR)*, *19*(2), 53-75.

Graham, P. A. (1987). Black teachers: A drastically scarce resource. *The Phi Delta Kappan*, *68*(8), 598-605.

Gregory, A., Hafen, C. A., Ruzek, E., Mikami, A. Y., Allen, J. P., & Pianta, R. C. (2016). Closing the racial discipline gap in classrooms by changing teacher practice. *School Psychology Review*, *45*(2), 171-191.

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

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

Heckman, J. J., & Snyder, J. M. (1996). *Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators.* NBER Working Paper No. 5785.

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

Hill, H. C. (2020, March). *Culturally responsive teaching is promising. But there's a pressing need for more research.* Education Week. *https://www.edweek.org/leadership/opinion-culturally-responsive-teaching-is-promising-but-theres-a-pressing-need-for-more-research/2020/03*

Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher*, *41*(2), 56-64.

Ho, Andrew D., and Thomas J. Kane. (2013). The reliability of classroom observations by school personnel. *Bill & Melinda Gates Foundation*.

Holt, S. B., & Gershenson, S. (2019). The impact of demographic representation on absences and suspensions. *Policy Studies Journal*, *47*(4), 1069-1099.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, *15*(4), 309.

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.

Irvine, J. J. (1989). Beyond role models: An examination of cultural influences on the pedagogical perspectives of Black teachers. *Peabody Journal of Education, 66*(4), 51-63.

Irvine, J. J., & Armento, B. J. (2001). Culturally responsive teaching: Lesson planning for elementary and middle grades. *Education Review*.

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.

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

Jencks, C., & Phillips, M. (1998). *The Black-White test score gap*. Brookings Inst Press.

Jensen, B., Grajeda, S., & Haertel, E. (2018). Measuring cultural dimensions of classroom interactions. *Educational Assessment, 23*(4), 250-276.

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

Kalogrides, D., Loeb, S., & Béteille, T. (2013). Systematic sorting: Teacher characteristics and class assignments. *Sociology of Education*, *86*(2), 103-123.

Kane, T. J., & Staiger, D. O. (2012). Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains. Research Paper. 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.

Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. *Review of Educational Research*, *86*(3), 643-680.

Kraft, M. A. (2019). Teacher effects on complex cognitive skills and social-emotional competencies. *Journal of Human Resources*, *54*(1), 1-36.

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.

Larson, K. E., Pas, E. T., Bradshaw, C. P., Rosenberg, M. S., & Day-Vines, N. L. (2018). Examining how proactive management and culturally responsive teaching relate to student behavior: Implications for measurement and practice. *School Psychology Review, 47*(2), 153-166.

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.

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

Liu, L., Hayes, M.S., & Gershenson, S. (2021). From referrals to suspensions: New evidence on racial discrimination in exclusionary discipline. *EdWorkingPaper No. 21-442*. Retrieved from: https://www.edworkingpapers.com/sites/default/files/ai21-442.pdf

Lyubomirsky, S., King, L., & Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success?. *Psychological Bulletin*, *131*(6), 803.

MacKinnon, D. P., Valente, M. J., & Gonzalez, O. (2020). The correspondence between causal and traditional mediation analysis: The link is the mediator by treatment interaction. *Prevention Science*, *21*(2), 147-157.

McCormick, M. P., Neuhaus, R., O'Connor, E. E., White, H. I., Horn, E. P., Harding, S., ... & McClowry, S. (2021). Long-term effects of social-emotional learning on academic skills: Evidence from a randomized trial of INSIGHTS. *Journal of Research on Educational Effectiveness*, *14*(1), 1-27.

Merolla, D. M. (2018). Completing the educational career: High school graduation, four-year college enrollment, and bachelor's degree completion among Black, Hispanic, and White students. *Sociology of Race and Ethnicity*, *4*(2), 281-297.

Mesler, R. M., Corbin, C. M., & Martin, B. H. (2021). Teacher mindset is associated with development of students' growth mindset. *Journal of Applied Developmental Psychology, 76*, 101299.

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

Mueller, G., & Plug, E. (2006). Estimating the effect of personality on male and female earnings. *ILR Review*, *60*(1), 3-22.

Murnane, R. J. (2013). US high school graduation rates: Patterns and explanations. *Journal of Economic Literature*, *51*(2), 370-422.

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

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

Perlman, M., Falenchuk, O., Fletcher, B., McMullen, E., Beyene, J., & Shah, P. S. (2016). A systematic review and meta-analysis of a measure of staff/child interaction quality (the classroom assessment scoring system) in early childhood education and care settings and child outcomes. *PloS one, 11*(12), e0167660.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*(2), 109-119.

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

Powell, R., Chambers Cantrell, S., Malo-Juvera, V., & Correll, P. (2016). Operationalizing culturally responsive instruction: Preliminary findings of CRIOP research. *Teachers College Record, 118*(1), 1-46.

Quillian, L., Pager, D., Hexel, O., & Midtbøen, A. H. (2017). Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. *Proceedings of the National Academy of Sciences*, *114*(41), 10870-10875.

Quinn, D. M. (2020). Experimental evidence on teachers' racial bias in student evaluation: The role of grading scales. *Educational Evaluation and Policy Analysis*, *42*(3), 375-392.

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.

Reardon, S. F., & Galindo, C. (2009). The Hispanic-White achievement gap in math and reading in the elementary grades. *American Educational Research Journal*, *46*(3), 853-891.

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.

Santamaria, L. J. (2009). Culturally responsive differentiated instruction: Narrowing gaps between best pedagogical practices benefiting all learners. *Teachers College Record, 111*(1), 214-247.

Sawchuk, S. (2021, May). *What is Critical Race Theory and why is it under attack?* Education Week. https://www.edweek.org/leadership/what-is-critical-race-theory-and-why-is-it-under-attack/2021/05

Shirrell, M., Bristol, T. J., & Britton, T. A. (2021). The effects of student-teacher ethnoracial matching on exclusionary discipline for Asian American, Black, and Latinx Students: Evidence from New York City. *EdWorkingPaper No. 21-475*. Retrieved from: https://edworkingpapers.org/sites/default/files/ai21-475.pdf

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.

Steinberg, M. P., & Garrett, R. (2016). Classroom composition and measured teacher performance: What do teacher observation scores really measure?. *Educational Evaluation and Policy Analysis*, *38*(2), 293-317.

Stronge, J. H. (2018). *Qualities of effective teachers*. ASCD.

Thompson, J., & Byrnes, D. (2011). A more diverse circle of friends. *Multicultural Perspectives, 13*(2), 93-99.

Tyson, W., Lee, R., Borman, K. M., & Hanson, M. A. (2007). Science, technology, engineering, and mathematics (STEM) pathways: High school science and math coursework and postsecondary degree attainment. *Journal of Education for Students Placed at Risk*, *12*(3), 243-270.

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

Valencia, Richard. (1997). *The evolution of deficit thinking*. London, United Kingdom: Falmer.

Valeri, L., & VanderWeele, T. J. (2013). Mediation analysis allowing for exposure–mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods, 18*(2), 137-150.

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.

Vincent, C. G., Randall, C., Cartledge, G., Tobin, T. J., & Swain-Bradway, J. (2011). Toward a conceptual integration of cultural responsiveness and schoolwide positive behavior support. *Journal of Positive Behavior Interventions, 13*(4), 219-229.

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.

Waters, M. M. (1989, September). An agenda for educating Black teachers. In The *Educational Forum* (Vol. 53, No. 3, pp. 267-279). Taylor & Francis Group.

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.

Yeager, D. S., & Dweck, C. S. (2020). What can be learned from growth mindset controversies?. *American Psychologist, 75*(9), 1269.

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., ... & Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, *573*(7774), 364-369.

Tables

Table 1

Demographic Characteristics of Teachers and Students in the Experimental Sample Relative to Larger Populations

	Experimental Sample	Full Project Sample		Full District Population	
Teacher Characteristics					
Female	0.85	0.78			
Asian	0.04	0.03			
Black	0.23	0.22			
Hispanic	0.03	0.03			
White	0.70	0.65			
Masters Degree	0.78	0.76			
Traditionally Certified	0.92	0.85	~		
Alternatively Certified	0.05	0.08			
No Formal Training	0.03	0.07			
Teaching Experience (years)	11.05	10.59			
Teacher Effects on State Math Test (standardized)	-0.004	0.016	***	0.001	
Teachers	71	321		3,559	
Student Characteristics					
Female	0.47	0.50	~	0.50	~
Asian	0.10	0.08	*	0.09	
Black	0.41	0.40		0.36	***
Hispanic	0.24	0.23		0.27	*
White	0.20	0.24	***	0.25	***
Other Race	0.04	0.04		0.05	
Free or Reduced-Price Lunch	0.68	0.64	**	0.63	***
Special Education	0.07	0.11	***	0.14	***
Limited English Proficiency	0.18	0.20		0.19	
Math Achievement (standardized)	0.038	0.101	*	0.008	
ELA Achievement (standardized)	0.050	0.084		0.006	
Students	1,283	10,586		175,572	

Notes: *** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1 on differences between the experimental sample and the full project sample or the full district populations.

Table 2
Descriptive Statistics for Student Outcomes

	Univariate Statistics					Pairwise Correlations between Outcomes in Upper-Elem. Sch.					
	Reliability	Upper Elem. Sch.		High Sch.		Self-Efficacy	Engage.	Self-Reg.	Math	ELA	Chron. Abs.
		Mean	SD	Mean	SD						
Intrapersonal Competencies											
Self-Efficacy (1 to 5)	0.76	4.09	0.64	NA		1					
Engagement/Happiness in Class (1 to 5)	0.82	3.99	0.90	NA		0.65***	1				
Self-Regulation (1 to 5)	0.74	4.09	0.93	NA		0.32***	0.24***	1			
Test Scores											
Math Achievement (standardized)	>0.90	0.04	0.92	0.06	0.92	0.31***	0.22***	0.22***	1		
ELA Achievement (standardized)	>0.90	0.05	0.90	0.19	0.85	0.25***	0.15***	0.28***	0.66***	1	
School Behaviors											
Chronically Absent (binary)	NA	0.05	NA	0.17	NA	-0.07*	-0.09**	-0.01	-0.13**	-0.06~	1
Ever Suspended (binary)	NA	0.04	NA	0.18	NA	-0.09**	-0.07*	-0.18**	-0.11**	-0.11*	0.05~

Notes: Reliability statistics for survey measures are internal consistency reliabilities; reliability statistics for math and ELA tests vary across participating school districts and grade levels, but all are above 0.9. Student-reported interpersonal competencies were not collected in high school. *** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1

Table 3

Descriptive Statistics for Teacher Mindsets and Practices

	Reliability	Means		Pooled SD	Regression-Adjusted Diff. (SD)	Pairwise Correlations			
		Black	White			Growth Mindset	Prep. for Inst.	Relat. Stu./Fam.	Class. Support
Growth Mindset Beliefs (1 to 6)	0.82	4.75	4.35	0.88	0.27***	1			
Preparation for Instruction (1 to 5)	0.78	3.36	3.23	0.36	0.13*	0.21~	1		
Relationships with Students and Families (1 to 5)	0.63	4.38	3.94	0.68	0.19**	0.23~	0.27*	1	
Classroom Support (1 to 7)	0.90; 0.47	4.43	4.15	0.38	0.19**	0.26*	-0.03	0.40**	1
Classroom Organization (1 to 7)	0.72; 0.63	6.60	6.42	0.27	0.19***	0.32**	0.21~	0.33**	0.21~

Notes: All measures have internal consistency reliability statistics; classroom observation scores also have adjusted intra-class correlations (listed second). Means for Black and White teachers and pooled standard deviations are reported on raw scales. Standardized differences between Black and White teachers are estimated from regression models that include school and grade fixed effects. *** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1

Table 4

Baseline Balance Tests

Student Characteristics	Stu. Charac. for Black Teachers	Diff. for White Teachers	Relationship Between Pre-Experimental Teacher Mindsets and Practices and Student Characteristics				
			Growth Mindset	Prep. for Instruction	Relat. Stu./Fam.	Classroom Support	Classroom Org.
Female	0.507	-0.005	0.007	0.007	0.005	0.003	0.017
		(0.012)	(0.022)	(0.016)	(0.023)	(0.022)	(0.013)
Asian	0.124	-0.053	-0.232	0.095	0.326**	-0.089	0.106
		(0.039)	(0.149)	(0.122)	(0.105)	(0.117)	(0.072)
Black	0.343	0.011	-0.130	0.014	0.091	-0.016	0.069
		(0.030)	(0.113)	(0.080)	(0.075)	(0.074)	(0.046)
Hispanic	0.247	-0.010	-0.099	0.084	0.171**	-0.010	0.070
		(0.026)	(0.115)	(0.079)	(0.064)	(0.080)	(0.051)
White	0.244	-0.044	-0.120	0.037	0.091	-0.039	0.085
		(0.029)	(0.119)	(0.105)	(0.076)	(0.100)	(0.056)
Free or Reduced-Price Lunch	0.676	0.002	-0.026	0.051	0.043	-0.026	-0.033
		(0.010)	(0.034)	(0.043)	(0.033)	(0.031)	(0.031)
Special Education	0.062	0.026	-0.062	0.116	-0.050	-0.027	-0.011
		(0.039)	(0.099)	(0.093)	(0.079)	(0.060)	(0.026)
Limited English Proficiency	0.216	-0.027	-0.079	-0.034	-0.130*	0.095	-0.039
		(0.026)	(0.065)	(0.048)	(0.057)	(0.061)	(0.032)
Prior Math Achievement	0.039	0.011	-0.018	0.004	-0.041	0.006	-0.014
		(0.016)	(0.033)	(0.029)	(0.030)	(0.027)	(0.017)
Prior ELA Achievement	0.032	0.000	0.019	0.009	-0.001	0.012	-0.018
		(0.011)	(0.033)	(0.035)	(0.023)	(0.030)	(0.017)
Prior Chronic Absenteeism	0.053	-0.010	-0.029	-0.040	-0.037	-0.070	0.007
		(0.045)	(0.109)	(0.084)	(0.091)	(0.114)	(0.040)
Prior Ever Suspended	0.026	-0.012	-0.206	0.243**	0.106	0.183	-0.140~
		(0.059)	(0.136)	(0.082)	(0.117)	(0.148)	(0.073)
P-values on Joint Tests of Significance:							
Estimates in each column		0.338	0.155	0.102	0.087	0.341	0.367
Estimates in all columns					0.390		
Teachers		71	71	71	71	71	71
Students		1,283	1,283	1,283	1,283	1,283	1,283

Notes: Average student characteristics of Black teachers are calculated from regression models that predict each characteristics as a function of race/ethnicity dummies for students' randomly assigned teacher, and school and grade fixed effects. Differences in average student characteristics for White teachers are calculated from a single linear probability regression model that predicts a dummy variable for the teacher being White as function of the student characteristics, a dummy indicator for whether or not a teacher is non-Black/non-White, and school and grade fixed effects. Regression models that predict pre-experimental measures of teacher mindset and practices as a function of student characteristics also control for school and grade fixed effects. Robust standard errors clustered at the teacher level in parentheses. *** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1

Table 5
Total Effects of Black versus White Teachers on Student Outcomes

Outcome	Sample Size	Average Effects Across All Students		Subgroup Effects		
				Black Students	Non-Black Students	P-Value on Diff.
	Panel A: End-of-Year Effects in Upper-Elementary School					
Self-Efficacy	903	0.364**	0.340**	0.563***	0.182	0.007
		(0.108)	(0.104)	(0.142)	(0.111)	
Engagement/Happiness	903	0.269*	0.242*	0.655***	-0.061	0.000
		(0.119)	(0.110)	(0.138)	(0.131)	
Self-Regulation	901	0.083	0.043	0.208	-0.084	0.119
		(0.109)	(0.102)	(0.159)	(0.117)	
Math Achievement	1,136	0.108	0.116*	0.112	0.115*	0.971
		(0.069)	(0.051)	(0.083)	(0.056)	
ELA Achievement	1,134	0.183*	0.197***	0.247**	0.164**	0.363
		(0.074)	(0.055)	(0.088)	(0.056)	
Chronically Absent	1,017	-0.045~	-0.044*	-0.046	-0.043~	0.915
		(0.025)	(0.021)	(0.029)	(0.023)	
Ever Suspended	1,017	0.055*	0.059*	0.049	0.068*	0.612
		(0.026)	(0.028)	(0.038)	(0.030)	
	Panel B: Follow-Up Effects in High School					
Math Achievement	663	0.033	0.004	0.017	-0.000	0.892
		(0.089)	(0.091)	(0.119)	(0.099)	
ELA Achievement	670	0.173*	0.155**	0.005	0.274***	0.021
		(0.073)	(0.058)	(0.084)	(0.075)	
Chronically Absent	743	-0.045	-0.051	-0.118*	0.000	0.060
		(0.040)	(0.039)	(0.055)	(0.045)	
Ever Suspended	743	-0.016	-0.009	0.029	-0.038	0.375
		(0.028)	(0.028)	(0.045)	(0.045)	
School and Grade Fixed Effects		X	X	X		
Background Student Characteristics			X	X		

Notes: For average total effects, 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 Black or non-Black/non-White (not shown), with White as the left-out category; and school and grade fixed effects. The first column of estimated total effects exclude background student characteristics, while the remaining models add student-level controls (i.e., gender, race/ethnicity, eligibility for free or reduced-price lunch, eligibility to receive special education services, limited English proficiency status, prior-year test scores in math and ELA, and prior-year absences and suspensions). For race-matching effects, estimates in each row come from a separate regression model that fully interacts teacher race/ethnicity categories and student background characteristics with student race/ethnicity dummies. Chronically absent and ever suspended variables are binary; other outcomes are standardized. Robust standard errors clustered at the teacher level in parentheses. *** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1

Table 6
Mediated Effects of Black versus White Teachers on Student Outcomes, on Average Across All Students

Outcome	Sample Size	Total Effect	% Change in Total Effect After Including Mediator(s)					
			All	Add Mediators One At a Time				
				Growth Mindset	Prep. for Inst.	Relat. Stu./Fam.	Class Support	Class Org.
Panel A: End-of-Year Effects in Upper-Elementary School								
Self-Efficacy	903	0.340**	-0.9	1.2	-0.3	-0.6	0.3	-0.6
		(0.104)						
Engagement/Hap.	903	0.242*	**-6.6**	**-6.6**	-3.3	-3.7	-0.8	-3.3
		(0.110)						
Self-Regulation	901	0.043	**-55.8**	9.3	7.0	16.3	4.7	**-48.8**
		(0.102)						
Math Achievement	1,136	0.116*	**-16.4**	0.0	-1.7	6.0	-4.3	**-23.3**
		(0.051)						
ELA Achievement	1,134	0.197***	4.1	-1.5	-1.0	5.1	4.1	3.0
		(0.055)						
Chronically Absent	1,017	-0.044*	0.0	6.8	2.3	-4.5	0.0	0.0
		(0.021)						
Ever Suspended	1,017	0.059*	1.7	**8.5**	1.7	-3.4	0.0	-1.7
		(0.028)						
Panel B: Follow-Up Effects in High School								
Math Achievement	663	0.004						
		(0.091)						
ELA Achievement	670	0.155**	1.9	-0.6	1.9	0.0	5.2	3.2
		(0.058)						
Chronically Absent	743	-0.051	**-11.8**	**-11.8**	-2.0	0.0	-3.9	-3.9
		(0.039)						
Ever Suspended	743	-0.009						
		(0.028)						

Notes: Total effects without mediators are the same as those listed in Table 5, from models that include background student characteristics in addition to school and grade fixed effects. The percent change estimates are calculated by dividing the direct effect of Black versus White teachers on student outcomes in models that include mediator(s) (not shown) minus the total effect, divided by the total effect. Rows are shaded in gray in instances where the total effect of Black versus White teachers is statistically significant. Percent change estimates are bolded when the magnitude is greater than five and when the mediator is associated with improved student outcomes. Percent change estimates are excluded when the total effect is very close to zero because dividing by a value close to zero explodes the percent change estimate. *** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1

Appendices

Appendix Table 1
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.
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".

Appendix Table 2

Pairwise Correlations between Upper-Elementary and High School Outcomes

Outcomes in Upper Elementary School	Outcomes in High School			
	Math	ELA	Chron. Abs.	Ever Susp.
Intrapersonal Competencies				
Self-Efficacy	0.26***	0.20***	-0.07~	-0.06
Engagement/Happiness	0.18***	0.15**	-0.08*	-0.05
Self-Regulation	0.22***	0.27***	-0.11**	-0.16***
Test Scores				
Math Achievement	0.69***	0.56***	-0.14**	-0.12**
ELA Achievement	0.55***	0.64***	-0.13***	-0.11**
School Behaviors				
Chronically Absent	-0.10**	-0.09*	0.32***	0.03
Ever Suspended	-0.13**	-0.13**	0.16***	0.19***

*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1

Appendix Table 3
Item Text from Teacher Survey

Dimensions and Item Text	Scale
<u>Growth Mindset Beliefs</u>	
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.	
<u>Preparation for Instruction</u>	
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.	
<u>Relationships with Students and Families</u>	
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.	

Appendix Table 4
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	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.

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

Appendix Table 5
Missing Data and Differential Attrition

	Students					Teachers (All Data)
	Upper-Elementary School			High School		
	Surveys	Test Scores	School Behaviors	Test Scores	School Behaviors	
Proportion with Data	0.70	0.89	0.79	0.52	0.58	0.86
Difference for White versus Black Teachers	-0.020	-0.010	0.046	0.058	0.062	0.059
	(0.062)	(0.020)	(0.051)	(0.048)	(0.046)	(0.103)
P-values on Differences in Background Characteristics	0.187	0.726	0.220	0.277	0.671	0.357
Observations	1,283	1,283	1,283	1,283	1,283	71

Notes: Differences in the proportion of students or teachers with data are estimated from separate linear probability regression models that predict a dummy indicator for whether or not a student or teacher had available data as a function of dummy indicators for the race/ethnicity of the teacher, as well as fixed effects for school and grade. None of the estimates is statistically significantly different from zero. P-values on differences in observable background characteristics come from linear probability regression models that predict whether or not students' randomly assigned teacher/the teacher was Black as a function of interactions between having the specific data source and all background characteristics of the student or teacher (see Table 1), as well as their main effects; a dummy indicator for whether or not students' randomly assigned teacher/the teacher was non-Black/non-White (with White teachers as the left-out category); and school and grade fixed effects. P-values come from a joint test of significance that all of the interaction terms are simultaneously equal to zero.

Appendix Table 6

Mediating Effects of Teacher Mindsets and Practices on Student Outcomes

	Self-Efficacy	Engage./ Hap.	Self-Reg.	Math Achieve.	ELA Achieve.	Chronic. Absent	Ever Susp.
	Panel A: End-of-Year Effects in Upper-Elementary School						
Black Teacher	0.337**	0.226~	0.019	0.097*	0.205***	-0.044*	0.060*
	(0.104)	(0.116)	(0.099)	(0.048)	(0.050)	(0.022)	(0.027)
Growth Mindset Beliefs	-0.020	0.029	-0.022	-0.025	0.005	0.005	-0.016*
	(0.053)	(0.062)	(0.044)	(0.028)	(0.024)	(0.007)	(0.008)
Prep. for Instruction	0.025	0.120	-0.007	0.055	0.068~	0.011	-0.017
	(0.060)	(0.078)	(0.050)	(0.038)	(0.039)	(0.012)	(0.012)
Relationships with Stu./Fam.	-0.001	-0.013	-0.034	-0.077*	-0.074**	-0.018*	0.014~
	(0.043)	(0.045)	(0.047)	(0.029)	(0.024)	(0.009)	(0.008)
Classroom Support	-0.016	0.013	-0.075~	0.004	-0.050*	-0.003	-0.001
	(0.035)	(0.056)	(0.041)	(0.025)	(0.021)	(0.007)	(0.009)
Classroom Organization	0.026	0.007	0.128*	0.161***	0.027	0.006	0.005
	(0.058)	(0.090)	(0.062)	(0.039)	(0.033)	(0.010)	(0.013)
Students	903	903	901	1,136	1,134	1,017	1,017
	Panel B: Follow-Up Effects in High School						
Black Teacher				-0.026	0.158**	-0.045	-0.007
				(0.085)	(0.057)	(0.037)	(0.026)
Growth Minset Beliefs				-0.031	0.008	-0.021	-0.036**
				(0.035)	(0.033)	(0.014)	(0.013)
Prep. for Instruction				-0.041	-0.030	-0.010	-0.004
				(0.049)	(0.054)	(0.019)	(0.024)
Relationships with Stu./Fam.				0.060	0.021	0.006	-0.009
				(0.037)	(0.035)	(0.013)	(0.017)
Classroom Support				-0.062	-0.053~	-0.017	0.001
				(0.037)	(0.031)	(0.013)	(0.016)
Classroom Organization				0.181***	0.006	0.010	0.040~
				(0.051)	(0.057)	(0.026)	(0.024)
Students				663	670	743	743

Notes: Estimates in each panel and column come from separate regression models that predict each student outcome as a function of whether or not students' randomly assigned teacher was Black or non-Black/non-White (not shown), with White as the left-out category; the full set of mediators; student-level controls (i.e., gender, race/ethnicity, eligibility for free or reduced-price lunch, eligibility to receive special education services, limited English proficiency status, prior-year test scores in math and ELA, and prior-year absences and suspensions); and school and grade fixed effects. Chronically absent and ever suspended variables are binary; other outcomes and teacher-level mindsets and practices are standardized. Robust standard errors clustered at the teacher level in parentheses. *** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1

Appendix Table 7

Mediated Effects of Black versus White Teachers on Student Outcomes, by Student Subgroup (Only for Instances of Differential Total Effects Across Student Subgroups)

Outcome	Total Effect		% Change in Total Effect After Including Mediator(s)					
				Add Mediators One At a Time				
	Black Students	Non-Black Students	All	Growth Mindset	Prep. for Instruction	Relat. Stu./Fam.	Classroom Support	Classroom Org.
Panel A: End-of-Year Effects in Upper-Elementary School								
Self-Efficacy	0.563***		-1.4	0.5	0.0	-0.4	0.9	-0.4
	(0.142)							
Engagement/Happiness	0.655***		-4.7	-3.7	-3.7	-2.9	-2.0	-2.0
	(0.138)							
Panel B: Follow-Up Effects in High School								
ELA Achievement		0.274***	0.7	1.8	-0.4	0.4	0.0	0.7
		(0.075)						
Chronically Absent	-0.118*		**-8.5**	**-8.5**	**-5.1**	-2.5	**-5.9**	**-5.1**
	(0.055)							

Notes: Total effects without mediators are the same as those listed in Table 5, from models that include background student characteristics in addition to school and grade fixed effects; student race/ethnicity categories are fully interacted with teacher race/ethnicity categories and student background characteristics. Mediating pathways only are shown in instances where there are differential effects of Black versus White teachers on the outcomes of Black versus non-Black students, and only for subgroup effects that are statistically significantly different from zero. The percent change estimates are calculated by dividing the direct effect of Black versus White teachers on the outcomes of Black or non-Black students in models that include mediator(s) (not shown) minus the total effect, divided by the total effect. Percent change estimates are bolded when the magnitude is greater than five and when the mediator is associated with improved student outcomes. *** $p<0.001$, ** $p<0.01$, * $p<0.05$, ~ $p<0.1$