Educational Evaluation and Policy Analysis
March 2018, Vol. 40, No. 1, pp. 59–81
DOI: 10.3102/0162373717714056

The Effects of Teacher Match on Students' Academic Perceptions and Attitudes

Anna J. Egalite
North Carolina State University
Brian Kisida
University of Missouri

Using student survey data from six U.S. school districts, we estimate how assignment to a demographically similar teacher affects student reports of personal effort, happiness in class, feeling cared for and motivated by their teacher, the quality of student–teacher communication, and college aspirations. Relying on a classroom fixed-effects strategy, we show that students assigned to a teacher with similar demographic characteristics experience positive benefits in terms of these academic perceptions and attitudes. The most consistent benefits are among gender matches, and the largest benefits are demonstrated by the combination of gender and racial/ethnic matches. The effects of gender matches are largely consistent across elementary and middle school, while the most consistent effects from race matches occur in middle school.

Keywords: *demographic mismatch, student perceptions, teacher diversity, academic attitudes, student effort*

In the fall of 2014, students of color outnumbered White students for the first time in U.S. public K–12 classrooms, while the gender balance remained evenly split. Meanwhile, teachers remain overwhelmingly female and White. Despite widespread acknowledgment of the demographic disparities between public school students and the teachers that serve them, numerous policy levers have been largely ineffective at addressing this divide (Goldhaber, Theobald, & Tien, 2015). Concurrently, persistent achievement gaps between minority and White students have only modestly improved since the 1960s (Clotfelter, Ladd, & Vigdor, 2009; Hanushek, Ingram, & Kenyon, 2014; Rampey, Dion, & Donahue, 2009). These troubling disparities are also observed in high school graduation rates, college enrollment, degree completion, and labor market outcomes (U.S. Department of Education, 2015). Additional gaps exist across gender lines, with female students often outperforming male students in reading, and male students often outperforming females in science and math (Dee, 2007; Robinson & Lubienski, 2011).

A growing body of evidence demonstrates that students benefit when assigned to a demographically similar teacher, especially racial/ethnic minority students. Such research has found that student–teacher demographic congruence is related to gains in student achievement (e.g., Dee, 2004; Egalite, Kisida, & Winters, 2015); more favorable teacher perceptions of student engagement, performance, and ability (Dee, 2005, 2007; Gershenson, Holt, & Papageorge, 2016; Grissom & Redding, 2016; Ouazad, 2014); reductions in student absences and suspensions (Holt & Gershenson, 2015); and a lower probability of students dropping out of high school (Gershenson, Hart, Lindsay, & Papageorge, 2017). These findings have supported arguments that the so-called teacher diversity gap and the teacher gender gap likely contribute to disparities in academic performance, bolstering policy directives aimed at diversifying the teacher labor force (Boser, 2011; Cherng & Halpin, 2016; Goldhaber et al., 2015) and informing approaches to teacher professional learning opportunities (Gay, 2010; Wallace, Kelcey, & Ruzek, 2016).

The specific mechanisms through which these benefits are realized, and how they may translate to long-term educational success and later-life outcomes, remain unclear. Commonly proposed theories about student–teacher demographic interactions tend to focus on the psychological and social effects that may occur when students are better able to view their teachers as role models (Boser, 2011; Evans, 1992; Zirkel, 2002), or when negatively biasing stereotypes of student–teacher interactions are abated (Ferguson, 1998). Such theories are generally grounded in the social and emotional aspects of student–teacher relationships from the student's perspective, yet most existing empirical studies have not been well-equipped to evaluate the dynamics of race/ethnicity and gender interactions through this lens. Fortunately, additional student measures are increasingly being collected, which have varyingly been referred to as noncognitive outcomes (Heckman & Rubinstein, 2001), social-emotional skills (Merrell & Gueldner, 2010), and academic behaviors and mind-sets (Blazar & Kraft, 2015). A growing body of evidence finds that teachers have measurable impacts on these types of outcomes (Blazar & Kraft, 2015; Gershenson, 2016; Jackson, 2012; Kraft & Grace, 2016). Yet with the notable exception of Holt and Gershenson's (2015) examination of the effect of student–teacher demographic mismatch on student attendance and suspensions and Gershenson and colleagues' (2017) examination of students' likelihood of dropout, noncognitive measures have not been thoroughly examined within the context of student–teacher demographic interactions (Grissom, Kern, & Rodriguez, 2015).

We address this gap in the literature by examining student self-reports of academic perceptions and attitudes (APA) that are directly tied to their classroom teachers. While earlier work (e.g., Dee, 2005; Gershenson et al., 2016; Ouazad, 2014) investigates match effects on teachers' perceptions of students, this article investigates match effects on students' perceptions of teachers, as well as assessments of classroom environment and self-reports of their academic engagement. Using data from surveys administered to more than 80,000 students as part of the Gates Foundation's Measures of Effective Teaching (MET) project, we are able to evaluate how gender and race/ethnicity interactions affect students' perceptions of these academic characteristics related to their teachers and classrooms. Seven of the measures collected are taken from Tripod surveys administered to students in Grades 4 to 8 (Ferguson, n.d.). The Tripod measures include scales indicating if a student feels cared for by his or her teacher (Care), student interest and enjoyment of classwork (Captivate), the quality of teacher–student communication (Confer), clarity in teaching style and methods (Clarify and Consolidate), students' self-assessment of their teachers' influence on their own effort and motivation (Effort), classroom management (Control), and students reporting if they feel pushed by their teachers (Challenge). We construct two additional attitudinal outcome scales from ancillary items collected by the MET researchers. These include a measure of students' happiness in class (Happy) and a measure of students' college aspirations (College).

Numerous theories of effective teaching emphasize a teacher's ability to motivate and provide social support (Ferguson & Danielson, 2014), as well as a teacher's classroom organization, instructional support, and emotional support (Pianta & Hamre, 2009). These characteristics are conceptually similar to the items included in the Tripod student perception survey. Prior research on Tripod survey items has found that they are reliable predictors of instructional quality (Kane & Cantrell, 2010; Kane & Staiger, 2012), and recent evidence finds that factors of responsivity and classroom management generated from Tripod items are significantly related to teacher value-added scores (Wallace et al., 2016). At the same time, however, related research has found that teachers who are most effective at improving test scores are not necessarily effective at improving students' behaviors and attitudes, supporting the theory that effective teaching is multidimensional (Blazar & Kraft, 2017; Jackson, 2016).

As such, these classroom-specific survey measures offer a unique window into the ways in which a student's classroom experience is affected by the persistent and widespread racial and gender teacher disconnect. We estimate race/ethnicity and gender interactions for these outcomes by exploiting the fact that each teacher is assigned ratings on these measures by multiple students. In previous studies, researchers have used similar student fixed-effects strategies to isolate the effect of student–teacher demographic interactions when students received multiple contemporaneous subjective ratings from different teachers (e.g., Dee, 2005; Gershenson et al., 2016; Papageorge, Gershenson, & Kang, 2016). In this case, because individual teachers are receiving ratings from multiple students, we reverse this intuition and use a classroom fixed-effects approach to isolate the effects of demographically similar teachers on student perceptions and academic attitudes. Moreover, because a subsample of students in the MET project was randomly assigned to teachers, we are able to provide some additional assurance that our strategy addresses the potential bias that could result from the nonrandom assignment of students to teachers.

Our results suggest there are important benefits for students' APA when they are assigned to a demographically congruent teacher. Using within-classroom comparisons, these effects are evident across both gender and racial/ethnic matches, with larger effects evident when students and teachers are demographically similar across both dimensions. The effects of gender matches are largely consistent across elementary and middle school students, while the most consistent effects from race matches occur in middle school, though there are some notable exceptions. Disaggregated results suggest that many of the largest benefits are demonstrated by White female students assigned to White female teachers, Black male students assigned to Black male teachers, and Black female students assigned to Black female teachers, compared with nonmatched students in the same classrooms.

Literature Review

Theoretical Framework

At least three distinct theories have been proposed to support calls for diversifying the teacher workforce to better serve students of color (Goldhaber et al., 2015). First, students may benefit from having a demographically similar teacher if they view their teachers as role models (Adair, 1984; Graham, 1987; Hess & Leal, 1997; Stewart, Meier, & England, 1989). In such instances, students may raise their academic motivations and aspirations when exposed to a demographically similar adult in a position of authority (King, 1993; Villegas, Strom, & Lucas, 2012). Exposure to a successful mentor could increase the cultural value that students ascribe to academic success, reduce the stigma of "acting White" (Fordham & Ogbu, 1986; Fryer & Torelli, 2010), and reduce instances of stereotype threat (Steele, 1997; Steele & Aronson, 1995), which occurs in situations where students feel pressure from a negative stereotype that inhibits their performance. Stereotype threat may be abated when teachers share their racial/ethnic or gender identity because teachers can affirm students' identity as one worthy of success and authority. Related research suggests that affirmation exercises abate stereotype threat (Cohen, Garcia, Apfel, & Master, 2006), though additional research has found that the effects may only emerge when moderated by supportive classroom environments (Dee, 2015). Furthermore, demographically similar teachers may serve to encourage students by adopting a mentoring role or advocating for students they identify with or who share backgrounds similar to their own (Adair, 1984; Graham, 1987; King, 1993; Ladson-Billings, 1992; Nixon & Robinson, 1999; Pitts, 2007; Stewart et al., 1989).

The second theory concerns the academic expectations that teachers hold for students, which prior research suggests are influenced by demographic similarities between students and teachers (Beady & Hansell, 1981; Ferguson, 2003; Gershenson et al., 2016; Grissom & Redding, 2016; Ouazad, 2014). If, as the evidence suggests, teachers' perceptions about student ability, aptitude, effort, and behavior are influenced by student race and/or gender, then students would benefit from increased exposure to teachers that are more representative of their students. Minority teachers may be more likely to push students to work hard and insist on a higher level of effort in class assignments.

The third theory concerns the potential for a deep and meaningful cultural understanding between teachers and students of similar backgrounds. Racially diverse teachers might be well positioned to design lessons that are culturally sensitive and to serve as “cultural translators.” Furthermore, if teachers are familiar with students’ cultural backgrounds, they might be less likely to succumb to unconscious biase stemming from negative stereotypes that alter the ways that teachers interact with students (e.g., Ferguson, 1998), especially if they hold stereotypes related to perceived academic ability (Rosenthal & Jacobson, 1968). Finally, a strong interpersonal connection between students and teachers stemming from a shared cultural understanding might reduce the likelihood of suspension, expulsion, or other extreme disciplinary response to student misbehavior, which has previously been shown to be susceptible to bias along racial or gender lines (Downey & Pribesh, 2004; Gregory, Skiba, & Noguera, 2010; Holt & Gershenson, 2015; Lindsay & Hart, 2017; McCarthy & Hoge, 1987).

Achievement Impacts

A number of studies have attempted to document student achievement benefits resulting from student–teacher pairings along race/ethnicity and gender lines. For example, an early study using a nationally representative dataset found no link between test score gains and same-race teachers (Ehrenberg, Goldhaber, & Brewer, 1995). In an analysis of data from Tennessee’s Project STAR class-size experiment, Dee (2004) found that third-grade Black and White students randomly assigned to racially similar teachers saw improved math and reading test scores by roughly 2 to 4 percentile points. Dee found the largest effects when Black students were assigned to Black teachers. Additional studies have found similar, though often smaller, effects using quasi-experimental approaches. For instance, Clotfelter, Ladd, and Vigdor (2006) and Goldhaber and Hansen (2010) use longitudinal data from North Carolina to document student achievement effects from racially congruent teachers of 0.02 to 0.03 *SD*. Using the same data, Goldhaber and Hansen find that Black students with Black teachers experienced the largest gains, at roughly 0.04 *SD*. Similarly, employing a student fixed-effects analysis with 8 years of data, Egalite et al. (2015) find some evidence of student–teacher matching effects in the range of 0.01 to 0.04 *SD*, with the strongest effects demonstrated by Black students in elementary grades.

Student achievement effects as a result of gender matches between teachers and students are less conclusive. Examining data from the National Education Longitudinal Study of 1988, Ehrenberg et al. (1995) find no evidence of achievement effects from student–teacher gender congruence. Analyzing data from the same source, however, Dee (2007) finds that assignment to an opposite-gender English teacher for 1 year reduces student achievement by 0.05 *SD*. Finally, Winters, Haight, Swaim, and Pickering (2013) analyze an administrative panel dataset from Florida and find no significant achievement impacts associated with student–teacher gender interactions. Rather, they conclude that both male and female students benefit from being assigned to a female teacher.

Effects on Nontested Academic Outcomes

Research has also examined student–teacher demographic congruence on subjective or “nontested” measures. Using the NELS:88 data, which include contemporaneous ratings of students by different teachers, Dee (2007) finds that assignment to a different gender teacher lowers teacher perceptions of student engagement and performance, with effect sizes ranging from −0.02 to −0.10 *SD*. Ehrenberg et al. (1995) reach similar conclusions using these same data. Ouazad (2014) also finds that students are rated stronger in terms of academic performance by same-race teachers. Similar to Dee (2005), Gershenson et al. (2016) exploit contemporaneous ratings by multiple teachers per student as an identification strategy. They find that non-Black teachers have lower expectations for the educational attainment of Black students, such that non-Black teachers are 12 percentage points more likely to expect Black students will only complete a high school diploma or less. A related paper by Gershenson et al. (2017) shows that exposure to one Black teacher in an elementary grade classroom reduces the probability of dropping out for Black students. Finally, Holt and Gershenson (2015) use a two-way fixed effects estimator to demonstrate the negative impact of student–teacher demographic mismatch on elementary students' absences and suspensions. They find that being assigned to a different race teacher leads to 0.04 more absences per year, and increases suspensions by 0.01 more times per year.

The Importance of Combining Research on Nontested Academic Outcomes and Teacher Effectiveness

Our research also speaks to emerging trends in teacher quality and its measurement. Of all the educational inputs within a school's control, none have been demonstrated to be as important as teachers (Hanushek, 2011; Winters, 2011). Based on measures of student achievement, having a higher quality teacher improves college attendance, leads to higher salaries, and lowers teen pregnancy rates (Chetty, Friedman, & Rockoff, 2014). Although the measurement of teacher quality has made significant progress, a narrow reliance on test score growth fails to capture what constitutes an effective teacher in nontested outcomes (Blazar & Kraft, 2015; Gershenson, 2016; Grissom, Loeb, & Doss, 2015; Jackson, 2016). For example, Kraft and Grace (2016), using related MET data, find substantial variation in teacher effects on students' social-emotional measures, such as self-regulation, growth mind-set, effort in class, and grit. Yet, similar to Blazar and Kraft (2017), they find only weak relationships between teachers who improve students' social-emotional measures and teachers who positively influence test scores, suggesting that quality teaching is multidimensional. The federal endorsement of broader measures of teacher quality in the recently adopted *Every Student Succeeds Act* (ESSA, 2015) is a particularly telling manifestation of the growing reluctance to rely on narrowly defined measures of teacher and school effectiveness, although some researchers are skeptical that newly adopted approaches have been thoroughly vetted (Duckworth & Yeager, 2015). Although the literature to date that attempts to demonstrate teachers' impact on students' noncognitive outcomes is nascent (Jennings & DiPrete, 2010; Ruzek, Domina, Conley, Duncan, & Karabenick, 2015), growing evidence validates the notion that skills and competencies other than standardized test performance can predict long-term outcomes (e.g., Almlund, Duckworth, Heckman, & Kautz, 2011; Heckman & Rubinstein, 2001; Tough, 2012). Our examination of these types of measures within the context of student–teacher demographic interactions adds to this emerging area of research, as student reports of teacher characteristics are a growing component of measures of teacher quality.

Data and Measures

The MET Project

Our data are drawn from the MET project, funded by the Bill and Melinda Gates Foundation, which tracked approximately 3,000 teachers in six school districts across the United States over the 2009–2010 and 2010–2011 school years. Those districts are Charlotte-Mecklenburg Schools (NC), the Dallas Independent School District (TX), Denver Public Schools (CO), Hillsborough County Public Schools (FL), Memphis City Schools (TN), and the New York City Department of Education (NY).

Over 2 years, MET researchers collected a variety of measures of teaching practice, including students' achievement on standardized tests, surveys of students' perceptions of their teacher and classroom environment, and videos of classroom practice. Seventy percent of the nearly 3,000 teachers in our sample participated in both years. Although the project relied on a volunteer sample of teachers in the six districts under study, the gender and racial characteristics of the teachers in the MET sample appear to reasonably reflect the districts these teachers represent. The primary difference is that volunteer teachers tended to have fewer years of teaching experience than the average for their districts (The Bill & Melinda Gates Foundation, 2010).

Description of the Randomization Procedure

In the second year of the study, the MET project team randomly assigned school-constructed classroom rosters to individual MET project teachers. To be included in the randomization sample, teachers had to share a school/grade/subject randomization block with at least one other MET project teacher. In all, 1,591 teachers were included in 668 randomization blocks in 284 participating schools. In an ideal scenario, schools would have complied with this random assignment, but a comparison of "assigned" teachers with "actual" teachers reveals substantial attrition from the study sample, with fidelity to the assigned teacher ranging from 27% in Memphis to 66% in Dallas (Kane, McCaffrey, Miller, & Staiger, 2013). Although we rely primarily on the strength of our classroom fixed-effects approach with statistical controls to identify the effects of students and teachers matched on gender and/or race, we also generate estimates using only the second-year randomly assigned subsample to provide additional assurance that our results are not a product of the nonrandom sorting of students to teachers. The noncompliance within the randomization procedure, however, tempers our ability to be absolutely certain that the effects we identify are causal. Throughout the article, we refer to this population as randomly assigned, as opposed to randomized, to reflect the noncompliance.

Sample and Descriptive Statistics

To construct our sample, we identify all students in Grades 4 through 8—those grades in which students took the state standardized assessment. We keep the three largest racial categories—White, Black, and Hispanic students—resulting in a final sample of 93,386 student observations. Because the MET project was designed to track teachers over time, and not students, we cannot follow students longitudinally. As a result, we are unable to know the extent to which students appear in multiple years, though the amount is likely substantial. Individual students can also appear more than once in the data if they are assigned to multiple teachers participating in the MET project in a given year. As a result, approximately 12% of the students each year enter the data multiple times, each time matched to a different teacher.[1]

Table 1 presents a descriptive overview of the sample. Twenty-eight percent of students are White, 38% are Black, and 34% are Hispanic. Due to small sample sizes, students who are not White, Black, or Hispanic are excluded from the analyses. More than half of students (57%) qualify for the federal free and reduced price lunch (FRL) program. Thirteen percent of students are classified as English language learners (ELL), 10% are gifted, and 9% are identified as having special educational needs. Teacher characteristics in our sample reflect the teacher racial diversity and teacher gender gaps seen nationwide. Just 18% of teachers are male, 57% are White, 37% are Black, and 6% are Hispanic. (Appendix Table A1, available in the online version of the journal, provides these student and teacher summary statistics broken out for each of the six districts under study.)

Table 1 also presents summary statistics for the 10 dependent variables, whose values range from 1 to 5. The full list of items included in each scale and the associated Cronbach's alphas are listed in Appendix B (available in the online version of the journal). Generally speaking, these scales have *SD*s of approximately 1, so even though we run our analysis using standardized versions of these variables, a 0.10 effect size can also be thought of as approximately a 0.10 point change on a 5-point Likert-type scale. Furthermore, Appendix Table C1, available in the online version of the journal, presents a matrix of correlation coefficients describing the strength and direction of the relationships between these various scales. The strongest observed relationship is between the Clarify and Care scales ($r = .71$). Similarly, Confer and Care are strongly related ($r = .70$). Conversely, the weakest observed relationship is between the Control and College scales ($r = .29$).

Table 2 provides the mean values for each of the 10 dependent variables broken out by various student characteristics. In general, it appears that some groups of students, on average, rate teachers higher or lower. Female students, for example, tend to give higher ratings than male students. The reasons for this difference could be numerous. One explanation may even be student/teacher matching (e.g., female students are more likely to be matched to female teachers because 82% of teachers in our sample are female). Although our main estimation strategy addresses this issue, the potential that some groups of students systematically give higher or lower ratings is an important point we return to when examining results using restricted models that do not allow us to control for this directly.

Our independent variables include "Same Race" and "Same Sex." We also include other possible combinations in subsequent models

TABLE 1

Descriptive Statistics of the Analytic Sample

Scale	*M*	*SD*	Minimum	Maximum	Observations
Student characteristics					
Male	0.50	0.50	0	1	93,386
White	0.28	0.45	0	1	93,386
Black	0.38	0.49	0	1	93,386
Hispanic	0.34	0.47	0	1	93,386
FRL	0.57	0.49	0	1	74,188
ELL	0.13	0.34	0	1	93,386
Gifted	0.10	0.30	0	1	93,386
Special education	0.09	0.28	0	1	92,788
Grade	6.09	1.37	4	8	93,386
Teacher characteristics					
Male	0.18	0.38	0	1	93,386
White	0.57	0.49	0	1	93,386
Black	0.37	0.48	0	1	93,386
Hispanic	0.06	0.23	0	1	93,386
Years of experience	10.01	8.83	0	46	42,509
Master's/advanced degree	0.36	0.48	0	1	76,007
Dependent variables					
Care	3.71	1.01	1	5	70,223
Captivate	3.59	0.95	1	5	70,425
Happy	3.83	1.07	1	5	69,875
Confer	3.68	0.86	1	5	70,419
Effort	4.01	0.71	1	5	70,298
College	3.77	1.18	1	5	56,605
Clarify	3.99	0.76	1	5	70,312
Control	3.42	0.84	1	5	70,283
Challenge	4.11	0.73	1	5	70,372
Consolidate	3.78	0.94	1	5	69,854

Note. n = 82,409 unique students, 1,909 teachers, 231 schools. FRL = free and reduced price lunch; ELL = English language learners.

(i.e., Same Race *and* Same Sex; Same Race *and* Other Sex; etc.). Table 3 provides a detailed breakdown of these independent variables by various student characteristics. Forty-eight percent of students are matched to a teacher of the same race, but this overall statistic masks important heterogeneity by race. White students are most likely to be matched to a same-race teacher at 80%; the corresponding statistic for Black students is 59%, and 9% for Hispanic students. All of these student-level percentages exceed the share of teachers' representation in our sample (57% of teachers are White, 37% are Black, and 6% are Hispanic), illustrating the dramatic nonrandom sorting of students and teachers across classrooms and schools. The aggregate statistic for assignment to a same-sex teacher (50%) also masks important heterogeneity by subgroups. Only 18% of male students are assigned to a male teacher, whereas 83% of female students are assigned to a female teacher.

Identification Strategy

The primary identification strategy builds upon that presented by Dee (2005) and Gershenson et al. (2016), which leverages contemporaneous subjective evaluations of students

TABLE 2

Unadjusted Means of the Dependent Variables, by Student Characteristics

Scale	All students	Male students	Female students	White students	Black students	Hispanic students	White male students	White female students	Black male students	Black female students
Care	3.71	3.67	3.75	3.68	3.77	3.66	3.63	3.73	3.74	3.81
Captivate	3.59	3.57	3.63	3.49	3.65	3.61	3.44	3.54	3.63	3.67
Happy	3.83	3.79	3.89	3.90	3.77	3.84	3.82	3.98	3.73	3.80
Confer	3.68	3.63	3.72	3.65	3.75	3.60	3.60	3.70	3.71	3.79
Effort	4.02	3.95	4.09	4.07	4.05	3.92	4.00	4.14	3.98	4.11
College	3.76	3.71	3.80	3.60	3.94	3.72	3.54	3.66	3.91	3.98
Clarify	3.98	3.94	4.02	3.94	4.04	3.96	3.92	3.97	3.99	4.08
Control	3.42	3.41	3.42	3.52	4.33	3.40	3.51	3.54	3.34	3.32
Challenge	4.11	4.08	4.14	4.09	4.18	4.06	4.06	4.12	4.15	4.22
Consolidate	3.78	3.75	3.80	3.64	3.87	3.80	3.63	3.64	3.84	3.90
n	82,409	41,344	41,065	22,138	32,505	27,766	11,203	10,935	16,133	16,372

Note. The standard deviations of the unstandardized dependent variables are as follows: Care: 1.01; Captivate: 0.95; Happy: 1.07; Confer: 0.86; Effort: 0.71; College: 1.18; Clarify: 0.76; Control: 0.84; Challenge: 0.73; Consolidate: 0.94.

TABLE 3
Sample Means of the Key Independent Variables, by Student Characteristics

	All	Male students	Female students	White students	Black students	Hispanic students	Grades 4–5	Grades 6–8
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Same race	0.48	0.48	0.48	0.80	0.59	0.09	0.52	0.45
Same sex	0.50	0.18	0.83	0.50	0.51	0.50	0.51	0.50
Same race, same sex	0.24	0.08	0.40	0.40	0.30	0.04	0.27	0.23
Same race, other sex	0.24	0.40	0.07	0.40	0.29	0.04	0.26	0.23
Other race, same sex	0.26	0.10	0.42	0.10	0.21	0.46	0.24	0.27
Other race, other sex	0.26	0.42	0.10	0.10	0.20	0.46	0.24	0.27

Note. n = 82,409 students, 1,909 teachers, 231 schools.

by teachers with various demographic characteristics. Because multiple students in our sample evaluate the same classroom teacher, we reverse this analytic approach to exploit the within-teacher variation from multiple students' evaluations. Formally, the measure of APA for student *i* in classroom *j* is specified as follows:

$$APA_{ij} = \beta_0 + \beta_1 \mathbf{Other}_i + \beta_2 \mathbf{X}_i + \alpha_j + \epsilon_{ij}, \quad (1)$$

where **Other** is a vector of variables that measure demographic mismatch between teacher and student. Following the convention established by Dee (2005), the **Other** vector in the baseline model contains two variables: *Other Race* and *Other Sex.* However, to test for multiplicative effects of assignment to a teacher who falls into both the *Other Race* and *Other Sex* categories, we follow Gershenson et al. (2016) and also consider a specification in which **Other** is composed of four mutually exclusive categories of demographic mismatch: *Same Race and Other Sex*, *Other Race and Same Sex*, *Other Race and Other Sex*, and *Same Race and Same Sex*, with the latter variable omitted as the reference category. We present results for both specifications. In terms of control variables in the model, **X** is a vector of observed student characteristics (i.e., gender, race, race-by-gender interactions, FRL, ELL, special education status, gifted status, and prior year math and reading test scores). These variables are intended to capture systematic differences in students' ratings that can be explained by observable characteristics, such as female students or Black students rating all teachers higher across the board. Similarly, if gifted students and students who score higher on standardized assessments have generally positive experiences with school, they might be more likely to rate teachers higher as part of a halo effect. Conversely, we also control for students who qualify for FRL[2] and ELL—two historically underperforming groups—in case these groups systematically assign more negative ratings to teachers because their overall educational experience is more negative. The most important control variable, however, is α, a classroom fixed effect that controls for unobserved classroom characteristics that might influence students' evaluations (for instance, the teacher's ability to motivate students in a given year). Finally, ϵ is a stochastic error term clustered at the school level.[3] Thus, β_1 is the coefficient of interest.

The key to our identification strategy is the classroom fixed effect. In previous studies, researchers have used similar student fixed-effects strategies to isolate the effect of student–teacher demographic interactions when students received multiple contemporaneous subjective ratings from different teachers (e.g., Dee, 2005; Gershenson et al., 2016). We reverse this intuition and exploit the fact that within each classroom-by-year, a teacher is assigned ratings by multiple students. As a result, the classroom fixed effect controls for the average ratings a teacher is assigned in a given year. This within-classroom estimation addresses the potential that teachers may have systematically higher or lower ratings that are related to time invariant characteristics, such as their race or gender. In addition, this approach accounts for any nonrandom sorting of students into classrooms based on unobserved factors, such as student motivation and ability. The classroom fixed effect also accounts for any unobserved time factors or anomalies such as annual variations in teacher quality, survey administration procedures, and other unobserved year specific factors that might influence student perceptions of their teacher. In addition, classroom fixed effects should largely eliminate reference group bias, which tends to hamper the ability to draw conclusions from student surveys across settings (West et al., 2016). As a result, this estimation approach allows us to precisely isolate within-teacher differences in students' subjective ratings that are systematically related to demographically matched or mismatched students.[4]

Finally, the key threat to internal validity with this identification strategy is the potential nonrandom sorting of students to teachers that systematically varies by student background within classrooms. In other words, if sorting mechanisms within classrooms are different for students of different genders or ethnicities, and these mechanisms are related to the gender or ethnicities of teachers, our identification of effects could be misleading. For example, if female students who rate teachers highly are systematically assigned to female teachers, and male teachers who rate teachers highly are systematically assigned to male teachers, then we could overestimate the size of the effect of student–teacher gender alignment. Although we note this caveat, we find no evidence to suggest that differential sorting is likely to pose a significant problem.[5]

Results

For our initial examination of the effect of teacher/student demographic match on students' APA, we estimate race and gender interactions separately by including indicator variables for *Other Sex* and *Other Race* (Table 4). We define the analysis sample in three different ways to ensure our findings are not influenced by sample characteristics. In columns 1 and 2, our estimates are generated using all available observations. Using this specification, we observe a consistent, statistically significant pattern of negative coefficients associated with the *Other Sex* variable for all but one of the scales examined, with significant effect sizes ranging from −0.02 to −0.06 *SD*. On average, students report having more favorable perceptions when their teacher is the same gender as them relative to students in the same class who do not share the gender of their teacher. There is no difference in the ratings for Control that are associated with the *Other Sex* variable.

Similarly, we observe statistically significant negative coefficients on the *Other Race* variable for Care, Clarify, and Control that range from −0.03 to −0.04 *SD*. On average, students whose teacher is the same race/ethnicity report having more favorable perceptions of their teacher across these three dimensions than students in the same classroom who do not share their teacher's race/ethnicity.[6]

Columns 3 and 4 add controls for prior math and English language arts (ELA) test scores. This restriction reduces the sample size due to some missing prior test scores in the data, but the substantive takeaways remain the same, with significant effect sizes ranging from −0.03 to −0.06 *SD*.

Finally, columns 5 and 6 report results from the second year of the MET study only, in which teachers were randomly assigned to class sections. The trade-off inherent in relying on this sample is one of statistical power versus eliminating some potential bias from the nonrandom sorting of students to class sections. Thus, we sacrifice a sample size of approximately 70,000 observations for one closer to 21,600. Nonetheless, the randomly assigned sample from the second year of the study allows us to be more confident that any inferences we draw about the relationship between teacher/student gender or racial congruence and the various outcomes examined are not driven by nonrandom student sorting. With this additional sample restriction, the direction of the effects is consistent with what we have observed thus far and, in some cases, the magnitude of the effects grows larger. In a few instances, however, the effects fall shy of statistical significance. Given the consistency of the direction and magnitude of the effects, the lack of statistical significance on some items is likely related to a reduction in study power when using this smaller sample.[7]

TABLE 4

Effects of Teacher/Student Demographic Match on Academic Perceptions and Attitudes

	All observations		Add prior test score control		Randomly assigned sample	
	Other race	Other sex	Other race	Other sex	Other race	Other sex
Scale	(1)	(2)	(3)	(4)	(5)	(6)
Care	−.03**	−.06***	−.03*	−.06***	−.03	−.06***
	(.01)	(.01)	(.02)	(.01)	(.02)	(.02)
	n = 69,852		*n* = 64,686		*n* = 21,575	
Captivate	−.01	−.06***	−.00	−.06***	−.04	−.05***
	(.01)	(.01)	(.01)	(.01)	(.03)	(.02)
	n = 70,054		*n* = 64,877		*n* = 21,602	
Happy	−.02	−.04***	−.02	−.04***	−.01	−.04*
	(.01)	(.01)	(.01)	(.01)	(.03)	(.03)
	n = 69,507		*n* = 64,407		*n* = 21,447	
Confer	−.02	−.04***	−.02	−.04***	−.01	−.03*
	(.01)	(.01)	(.01)	(.01)	(.02)	(.02)
	n = 70,048		*n* = 64,875		*n* = 21,594	
Effort	−.03*	−.02*	−.03	−.03**	−.01	−.05**
	(.02)	(.01)	(.02)	(.01)	(.03)	(.02)
	n = 69,927		*n* = 64,766		*n* = 21,587	
College	−.02	−.06***	−.02	−.06***	NA	NA
	(.02)	(.01)	(.02)	(.01)		
	n = 46,236		*n* = 42,851			
Clarify	−.03**	−.04***	−.03*	−.04***	−.02	−.02
	(.02)	(.01)	(.02)	(.01)	(.02)	(.02)
	n = 69,941		*n* = 64,774		*n* = 21,585	
Control	−.04***	−.01	−.04***	−.01	−.04*	−.03
	(.01)	(.01)	(.01)	(.01)	(.02)	(.02)
	n = 69,912		*n* = 64,758		*n* = 21,575	
Challenge	−.01	−.02**	−.01	−.03***	.02	−.03
	(.01)	(.01)	(.02)	(.01)	(.02)	(.02)
	n = 70,001		*n* = 64,836		*n* = 21,600	
Consolidate	−.02	−.03***	−.02	−.03***	−.01	−.05**
	(.01)	(.01)	(.02)	(.01)	(.03)	(.02)
	n = 69,488		*n* = 64,378		*n* = 21,486	
Prior test scores	No		Yes		Yes	

Note. Models include controls for student gender, student race, the interactions of student gender by student race, FRL, ELL, special education, gifted, prior year math and ELA scores (test scores included in columns 3 through 6 only), classroom fixed effects, and a missing data indicator variable for FRL. Standard errors in parentheses are robust to clustering at the school level. FRL = free and reduced price lunch; ELL = English language learners; ELA = English language arts.
**p* < .1. ***p* < .05. ****p* < .01.

In summary, across all three specifications, the largest and most consistently significant effects are observed for Care, Captivate, Happy, Confer, Effort, and Consolidate. In general, the effects are larger and more consistent among gender matches.[8]

In Table 5, we report estimates from a more finely specified version of the analytical model in which the **Other** vector is specified as a set of four mutually exclusive categories that describe the specific nature of the demographic match between students and teachers: *Same Race and*

TABLE 5

Effects of Teacher/Student Demographic Match on Academic Perceptions and Attitudes, Results by All Demographic Pairings

	All observations			Add prior test score control			Randomly assigned sample		
	SROS	ORSS	OROS	SROS	ORSS	OROS	SROS	ORSS	OROS
Scale	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Care	−.07***	−.04**	−.09***	−.07***	−.03*	−.09***	−.06**	−.03	−.09***
	(.01)	(.02)	(.02)	(.02)	(.02)	(.02)	(.03)	(.03)	(.03)
Captivate	−.08***	−.02	−.07***	−.08***	−.02	−.06***	−.08***	.02	−.02
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.03)	(.03)	(.03)
Happy	−.06***	−.03*	−.07***	−.06***	−.03*	−.06***	−.04	−.01	−.05
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.03)	(.03)	(.04)
Confer	−.04***	−.02	−.06***	−.05***	−.03	−.06***	−.06**	−.04	−.05*
	(.01)	(.02)	(.02)	(.01)	(.02)	(.02)	(.03)	(.03)	(.03)
Effort	−.03*	−.04**	−.05***	−.03**	−.03*	−.05**	−.05*	−.02	−.06*
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.03)	(.03)	(.04)
College	−.07***	−.03	−.08***	−.07***	−.03	−.08***	NA	NA	NA
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)			
Clarify	−.05***	−.04**	−.07***	−.06***	−.04**	−.07***	−.04	−.03	−.04
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.03)	(.03)	(.03)
Control	−.02	−.05***	−.05***	−.02	−.05***	−.06***	−.03	−.04	−.06**
	(.01)	(.02)	(.02)	(.01)	(.02)	(.02)	(.02)	(.03)	(.03)
Challenge	−.04***	−.03	−.04**	−.05***	−.03	−.04**	−.05*	.01	−.01
	(.01)	(.02)	(.02)	(.02)	(.02)	(.02)	(.03)	(.03)	(.03)
Consolidate	−.04**	−.03*	−.06***	−.04***	−.03	−.05***	−.05*	.01	−.04
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.03)	(.03)	(.03)
Prior test scores		No			Yes			Yes	

Note. Models include controls for student gender, student race, the interactions of student gender by student race, FRL, ELL, special education, gifted, prior year math and ELA scores (test scores included in columns 4 through 9 only), classroom fixed effects, and a missing data indicator variable for FRL. Standard errors in parentheses are robust to clustering at the school level. Specific sample sizes are identical to those reported in Table 4. In columns 1 to 3, n ranges from 46,236 to 70,054. In columns 4 to 6, n ranges from 42,851 to 64,877. In columns 7 to 9, n ranges from 21,447 to 21,602. SROS = same race, other sex; ORSS = other race, same sex; OROS = other race, other sex; FRL = free and reduced price lunch; ELL = English language learners; ELA = English language arts.

$^*p < .1$. $^{**}p < .05$. $^{***}p < .01$.

Other Sex, *Other Race and Same Sex*, *Other Race and Other Sex*, and *Same Race, Same Sex*. All effect sizes are judged relative to a *Same Race, Same Sex* match, which is the omitted category. Typically, the effects in columns 1, 4, and 7 are larger than columns 2, 5, and 8, confirming the patterns from Table 3. That is, the negative effects of gender mismatches are generally of higher magnitude than race/ethnicity mismatches. Specifically, the statistically significant coefficients on same race, other sex (SROS) range from −0.04 to −0.08 *SD*, whereas the coefficients on other race, same sex (ORSS) range from −0.04 to −0.05 *SD*. In some cases, the largest negative coefficients are observed in columns 3, 6, and 9, which are instances when a student experiences both a racial and gender mismatch (other race, other sex [OROS]). The statistically significant coefficients range from −0.04 to −0.09 *SD*. Finally, as we observed before, the results from the randomly assigned sample are well-aligned with the results that rely exclusively on classroom fixed effects.

Elementary and Middle School Subgroups

It is also instructive to break out the impacts on APA by grade level, as students at different developmental stages may benefit in different ways from assignment to a demographically similar teacher. Table 6 presents the results of this subgroup analysis. What stands out first is the consistency of statistically significant findings for gender matches (SROS) across the two grade groupings. Looking at the results when race matching is included (ORSS and OROS), however, the estimates are more consistently statistically significant for middle school students. Examples include Care—where the OROS coefficient is −0.05 *SD* for elementary students and −0.09 *SD* for middle school students—and Captivate, Happy, Effort, College, Challenge, and Consolidate—where the SROS coefficient is insignificant for elementary students and ranges from −0.06 to −0.10 *SD* for middle school students. Control, on the contrary, shows consistent significant effects in both the elementary and middle school samples. Both the ORSS and OROS coefficients are −0.07 and −0.05 *SD* for elementary and middle school students, respectively.

A particularly interesting finding is that middle school students matched on both race and gender are more likely to report thinking about college more because of their teacher. The OROS coefficient is −0.10 *SD*, whereas the corresponding statistic for elementary-aged students is not statistically significant. They may be too young to have serious college aspirations and are thus unaffected in this area when assigned to a demographically similar teacher.

Another notable difference is present in the Clarify variable. In this case, it is the elementary-aged students who are more likely to report more favorable ratings for same-sex and same-race teachers. Indeed, the −0.11 *SD* coefficient for elementary students on OROS is the largest statistically significant finding in Table 6. Similarly, Confer is larger for elementary students in the OROS column. This may reflect differences in students' developmental stages and their ability to independently seek and retrieve information related to their coursework. It is possible that younger students find themselves relying more on the teacher for explanation and clarification, thus leaving themselves more open to being impacted by a demographic match.

The most consistently large and negative results of this subgroup analysis are observed for middle school students experiencing the double-impact of a gender and race mismatch. In particular, the coefficients on the OROS variable for middle school students are negative and statistically significant across all 10 outcomes, with coefficients ranging from −0.04 to −0.09 *SD*. Thus, while both age groups appear to benefit from assignment to a demographically similar teacher, it is students in Grades 6 through 8 who experience the negative impact of a mismatch most consistently, and this seems to be driven by the additive effect of a combined gender and race mismatch.

Effects Disaggregated by Teacher Race and Gender

Thus far, we have restricted the effect of a demographic mismatch to be constant across all students, but this approach may mask important differences by race or gender. Table 7 reports results when the sample is restricted to White male teachers, White female teachers, Black

TABLE 6

Academic Perceptions and Attitudes, Results for Elementary and Middle School Student Subgroups

	Same race, other sex		Other race, same sex		Other race, other sex	
	Elementary grades (4–5)	Middle grades (6–8)	Elementary grades (4–5)	Middle grades (6–8)	Elementary grades (4–5)	Middle grades (6–8)
	(1)	(2)	(3)	(4)	(5)	(6)
Care	−.06**	−.06***	.00	−.04	−.05*	−.09***
	(.03)	(.02)	(.03)	(.02)	(.03)	(.02)
Captivate	−.05*	−.08***	−.00	−.04	−.02	−.07***
	(.03)	(.02)	(.03)	(.02)	(.03)	(.02)
Happy	−.04	−.06***	−.01	−.04*	−.02	−.07**
	(.03)	(.02)	(.03)	(.02)	(.03)	(.02)
Confer	−.04	−.05***	−.05	−.02	−.07**	−.04**
	(.03)	(.02)	(.03)	(.02)	(.03)	(.02)
Effort	−.03	−.03*	.00	−.04*	−.02	−.06**
	(.03)	(.02)	(.03)	(.02)	(.03)	(.02)
College	−.03	−.09***	.04	−.06*	−.03	−.10***
	(.04)	(.02)	(.04)	(.03)	(.04)	(.03)
Clarify	−.07**	−.06***	−.07**	−.02	−.11***	−.05**
	(.03)	(.02)	(.03)	(.02)	(.03)	(.02)
Control	−.01	−.03*	−.07**	−.05**	−.07**	−.05**
	(.03)	(.02)	(.03)	(.02)	(.03)	(.02)
Challenge	−.07**	−.05**	−.04	−.03	−.01	−.07***
	(.03)	(.02)	(.03)	(.02)	(.03)	(.02)
Consolidate	−.06**	−.05**	.01	−.05**	−.05	−.07***
	(.03)	(.02)	(.03)	(.02)	(.04)	(.02)

Note. Separate regressions were run for elementary and middle school grades. Columns 1, 3, and 5 come from the elementary grades regression ($n = 24{,}327$); columns 2, 4, and 6 come from the middle school grades regression ($n = 40{,}608$). Models include controls for student gender, student race, the interactions of student gender by student race, FRL, ELL, special education, gifted, prior year math and ELA scores, classroom fixed effects, and a missing data indicator variable for FRL. Standard errors in parentheses are robust to clustering at the school level. FRL = free and reduced price lunch; ELL = English language learners; ELA = English language arts.
$^{*}p < .1$. $^{**}p < .05$. $^{***}p < .01$.

male teachers, and Black female teachers. To provide the most reliable estimates, all estimates in Table 7 are based on the second-year sample, in which teachers were randomly assigned to classrooms, except in the case of College, which was only asked in the first year. Although we also generated estimates for Hispanic teachers, we found no strong evidence of effects from demographic matches; thus, we exclude Hispanic teachers and students from the table for ease of presentation.

To interpret Table 7, the reader should bear in mind that the omitted student category is always the same race and same sex as the teacher. For example, in the case of columns 1 through 3, which are restricted to White male teachers, all effects are relative to White male students. Thus, column 1 represents the ratings assigned to a White male teacher by a White female student, relative to the ratings assigned to a White male teacher by a White male student in the same class.

In addition, although the results in Table 7 more precisely illustrate the underlying patterns driving our main results, it is important to bear in mind that when we restrict the sample to teachers of a single gender and race in our classroom fixed-effects model, we can no longer control for the average effects of student gender and race because those controls are perfectly collinear with the match terms in the model. As such,

TABLE 7

Subgroup Estimates of the Effects of Teacher/Student Demographic Match on Academic Perceptions and Attitudes, Using Randomly Assigned Sample

	White male teachers			White female teachers			Black male teachers			Black female teachers		
	White female students	Black male students	Black female students	White male students	Black female students	Black male students	Black female students	White male students	White female students	Black male students	White female students	White male students
Scale	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Care	−.02	−.04	−.08	−.09***	−.02	−.09**	−.01	−.06	−.12	−.13***	−.08	−.13*
	(.09)	(.08)	(.07)	(.03)	(.03)	(.04)	(.09)	(.10)	(.11)	(.03)	(.06)	(.06)
Captivate	−.00	.13	.10	−.17***	.03	−.04	−.04	−.14	−.12	−.06*	−.10	−.15**
	(.09)	(.08)	(.08)	(.03)	(.04)	(.04)	(.10)	(.13)	(.13)	(.03)	(.07)	(06)
Happy	.08	.10	.06	−.20***	.01	−.08**	.07	.11	−.10	−.07	−.05	−.10
	(.08)	(.09)	(.09)	(.03)	(.03)	(.04)	(.09)	(.11)	(.12)	(.03)	(.07)	(.06)
Confer	.03	−.01	.04	−.08***	−.07*	−.02	−.05	−.24**	−.20*	−.17***	−.15**	−.22***
	(.07)	(.09)	(.08)	(.03)	(.04)	(.04)	(.11)	(.12)	(.12)	(.03)	(.06)	(.06)
Effort	.14*	−.14	.02	−.22***	−.02	−.24***	.06	−.23	−.07	−.18***	.00	−.21***
	(.08)	(.09)	(.10)	(.03)	(.04)	(.04)	(.10)	(.14)	(.16)	(.03)	(.06)	(.06)
College	−.00	.08	.01	−.12***	−.00	−.07**	−.00	.00	.09	−.12***	−.17***	−.28***
	(.06)	(.06)	(.06)	(.02)	(.03)	(.03)	(.05)	(.09)	(.09)	(.02)	(.05)	(.05)
Clarify	−.07	−.00	.08	−.03	.10***	−.01	−.10	.04	.09	−.15***	−.18***	−.24***
	(.09)	(.09)	(.09)	(.03)	(.04)	(.04)	(.09)	(.10)	(.12)	(.03)	(.06)	(.06)
Control	−.09	.05	−.03	−.01	.04	.08**	−.08	−.16	−.19	.03	−.13**	−.10*
	(.07)	(.08)	(.08)	(.03)	(.03)	(.03)	(.80)	(.10)	(.12)	(.03)	(.06)	(.06)
Challenge	−.05	.06	.06	−.09***	.04	−.03	.00	−.11	−.07	−.11***	−.10	−.17***
	(.08)	(.09)	(.08)	(.03)	(.03)	(.04)	(.09)	(.12)	(.12)	(.03)	(.06)	(.06)
Consolidate	−.06	.06	−.02	.02	.10***	.03	−.02	−.09	−.12	−.09***	−.18**	−.20***
	(.08)	(.08)	(.08)	(.03)	(.04)	(.04)	(.09)	(.12)	(.12)	(.03)	(.07)	(.06)

Note. Models include controls for student gender, student race, FRL, ELL, special education, gifted, prior year math and ELA scores, classroom fixed effects, and a missing data indicator variable for FRL. Standard errors in parentheses are robust to clustering at the school level. In columns 1 to 3, *n* ranges from 2,079 to 4,503. In columns 4 to 6, *n* ranges from 10,807 to 21,135. In columns 7 to 9, *n* ranges from 1,118 to 2,563. In columns 10 to 12, *n* ranges from 6,184 to 12,226; Outcomes for the College variable come for the first year of data collection, whereas all other outcomes come from the second-year, randomly assigned sample. FRL = free and reduced price lunch; ELL = English language learners; ELA = English language arts.
$^*p < .1$. $^{**}p < .05$. $^{***}p < .01$.

if certain types of students (e.g., female Black students) consistently give higher or lower ratings than other types of students, this could give the appearance of matching effects if examined in isolation. As a result, it is important to compare across column groupings before drawing strong conclusions.[9]

Broadly speaking, gender matches appear to be driven by White and Black male students rating female teachers, as judged by the number of significant results in columns 4 and 10.

The effects of race matching appear to be mostly driven by Black students rating Black teachers higher, given the size and significance of estimates in columns 8 and 11. Finally, the combined effect of gender and race matching yields some of largest and most consistent effects (columns 3, 6, 9, and 12), with some of the largest effects obtained when comparing Black female students to White male students in the same classroom led by Black female teachers (column 12). The "reverse" effects in column 3, which explores the differential between these two groups of students in the same classroom led by a White male teacher, do not suggest that the results are spurious. Large effects are also common in column 6, which compares White female and Black male students in the same classroom led by White female teachers, a result reinforced by the generally negative "reverse" effects in column 9. There are, however, a couple of instances of positive mismatch effects in column 5, illustrating instances where Black female students rate White female teachers higher than White female students on Confer and Clarify, yet these effects are smaller than instances where the situation is reversed and Black female students are rating Black female teachers relative to White female students (column 11).

In terms of particular outcome variables, few notable trends emerge. Most results for Care, Confer, Effort, College, Challenge, and Consolidate follow a general trend of being significant when matched students are compared with students of different gender and same race in the same classroom, with effects that tend to increase when matched students are compared with students of both different gender and different race in the same classroom.

Discussion

Using a classroom fixed-effects model, we identify the effect of student–teacher demographic matches on students' ratings across 10 APA. Across a number of different specifications, our findings demonstrate that students who share gender and/or racial characteristics with their teachers have more positive perceptions of their teachers in terms of feeling cared for, feeling that their schoolwork is interesting, and more positive reports of instructional characteristics related to student–teacher communication and guidance compared with unmatched students in the same classroom. They also report putting forth more personal effort and have higher college aspirations. This study provides important evidence that demographic matches influence students' APA and may shed light on the specific ways in which students are affected by the wide demographic divide between teachers and students in American public education. The largest effects tend to be concentrated when students share both gender and racial characteristics with their teachers, compared with students who share neither.

When we examine effects for elementary and middle school students separately, we see evidence of some heterogeneous effects. Middle school students who experience a race and gender match with their teacher are more likely to say that because of their teacher, they think more about going to college. For elementary students who experience a race and gender match, they are more likely to say that they can understand what they are supposed to be learning in class and that their teacher explains difficult things clearly (e.g., Confer and Clarify scales).

These results are particularly meaningful when considered in the context of the major demographic shift that has occurred in American public schools in recent years, which has tipped the balance toward a majority–minority student population. Between 1992 and 2012, the proportion of students who were White dropped from 67% to 51% whereas the proportion of students who were Hispanic rose from 12% to 24% and the proportion of students who were Black held steady at roughly 16% (Snyder, 2016). During this same time period, however, the teacher workforce remained overwhelmingly white and female.

Our findings generally relate to the theories that motivate calls to diversify the teacher labor force. In terms of "role modeling," significant effects on the College scale show that students assigned to demographically similar teachers think more about going to college because of their teacher compared with other students in the same classroom. This is consistent with prior theory and research, which suggests demographically similar teachers may be more likely to encourage students or serve as mentors to students with whom they share similar backgrounds (King, 1993). Against this backdrop, our findings on students' college aspirations make intuitive sense.

We also find that racial and gender similarities between students and teachers result in higher ratings on the Challenge and Effort scale. These measures capture student reports about their teacher pushing them to work hard, accepting nothing less than their full effort, and motivating them to do their best quality work. These outcomes are closely related to prior theory that teachers may hold higher expectations for demographically similar students. These effects appear to be most meaningful for female students, particularly for Black female students linked with Black female teachers. This is consistent with prior research, which has shown that Black teachers hold higher expectations for Black students (Fox, 2015; Gershenson et al., 2016; Ouazad, 2014).

What we find particularly compelling is the evidence we uncover in support of the theory of cultural understanding, which suggests that teachers of color may be particularly well situated to explain new material in a culturally relevant and engaging way. Our results offer evidence in support of this theory, and the effects are particularly strong for Black students paired with Black teachers. Compared with students in the same classroom, we see large effects on the Care, Confer, Clarify, and Consolidate scales for Black female students paired with Black female teachers, as well as a large effect for Black male students on the Confer scale (ORSS = −0.24 *SD*). These outcomes generally support the theory of cultural understanding and are related to perceived differences in instructional techniques. Items in the Confer, Consolidate, and Clarify scales measure student reports about how much their teacher asks questions to make sure they understand class material, explains what they are learning and why, explains things in a different way if they do not understand something, provides helpful comments about mistakes on assignments, and invites them to share insights and ideas. This is consistent with the theory that demographically similar teachers are well positioned to employ targeted instructional approaches (King, 1993), serve as cultural translators (Irvine, 2000; King, 1993), and employ "culturally relevant pedagogy" (Ladson-Billings, 1994, 1995). The "cultural understanding" theory also supports the idea that students of color assigned to diverse teachers might be more likely to feel cared for and happy in class, which in turn may motivate them to work hard and aspire high. The significant results we observe for the Care scale, which includes items such as "I like the way my teacher treats me when I need help" and "My teacher in this class makes me feel that he or she really cares about me," align nicely with this theory.

Our results and additional recent research (e.g., Gershenson et al., 2017; Holt & Gershenson, 2015) suggest that studies focused on achievement effects may have only observed the tip of the iceberg with regard to the benefits of demographically similar teachers. The findings presented here provide additional evidence for the potential benefits of policies that address the teacher diversity gap. At a micro level, these findings can inform the design of professional development opportunities for all teachers to address the differences in student perceptions based on mismatches in student–teacher characteristics. Efforts to educate teachers with the tools to engage in culturally responsive teaching may be a strategy to improve existing pedagogical practices in the short-term (Gay, 2010; Weinstein, Tomlinson-Clarke, & Curran, 2004). At a macro level, the results presented here provide evidence of the specific ways in which a student's classroom experience is affected by the persistent and widespread lack of racial and gender representation in the teacher labor force. For policymakers, this study provides strong support for innovative and bold actions to reduce barriers to entry for more diverse teachers entering the profession and efforts to improve retention.

It is important to acknowledge a number of limitations that apply to this work. First, the six school districts in this study are urban districts, and these same results may not hold in other locations. Second, the measures we rely on are generated from student reports, and thus the results might be driven by a bias for demographically similar teachers instead of reflecting substantial differences in actual instructional practices or classroom management. Only by validating these measures through external means could we fully answer this question. Future research needs to determine the extent to which students' reports of academic perceptions accurately reflect classroom practices. Moreover, future research must determine if students' self-reports such as the ones we examine translate into tangible benefits in school and later-life outcomes, such as high school and college attainment and employment. Such information will serve as a critical benchmark to help readers interpret the magnitude of the effects observed here. In light of our findings and the growing emphasis on multiple measures of academic success and teacher quality, researchers must make efforts to validate the importance of these types of measures in more concrete terms and use that information to contextualize the magnitude of the impacts we observe on these outcomes. Only then can the full implications of the teacher gender gap and the teacher diversity gap be fully considered.

Authors' Note

Anna J. Egalite and Brian Kisida contributed equally to the work presented here, and the authors' names appear in alphabetical order.

Acknowledgments

The authors thank Susan Dynarski, Raegan Miller, Seth Gershenson, Cory Koedel, Eric Parsons, and participants at the 2016 meeting of the Association for Education Finance and Policy, as well as participants at the 2017 Microeconomics Workshop Series at North Carolina State University for helpful feedback on earlier drafts of this article. This article also benefited greatly from the thoughtful and helpful critiques provided by the reviewers and editor.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. We conducted tests to ensure that this lack of independence was not substantively affecting our findings. First, we randomly dropped within-year duplicate observations (for students with different math and reading teachers) and reran our models with this smaller sample. The findings reveal point estimates that are qualitatively similar in direction, magnitude, and significance to our main results. We also ran the analysis on a single year of data (which solves the issue of students appearing in two years of data) and randomly drop duplicate records as above. As before, the findings reveal point estimates that are fundamentally similar in direction, magnitude, and significance to our main results.

2. Because free or reduced lunch status is not reported for one of the districts in our sample, we follow the convention of Garrett and Steinberg (2015) who use a missing data dummy variable when examining these same data to retain these observations. Our results when excluding this district are consistent with our main findings.

3. Clustering at the teacher level makes intuitive sense because this is the source of treatment variation in our analysis and is the most commonly adopted method in related research (Garrett & Steinberg, 2015; Kane, McCaffrey, Miller, & Staiger, 2013; Koedel, Parsons, Podgursky, & Ehlert, 2015), but our standard errors could be inflated under this approach because students are also clustered within schools. Moreover, a fraction of the students in our sample appear more than once in different classrooms with different teachers, and teacher-level clustering would fail to adjust for this. Thus, we adopt the advice of Angrist and Pischke (2009) and "pass the clustering buck one level higher" by clustering at the school level. Of note, the difference between clustering standard errors at the teacher level versus at the school level is miniscule. We also ran a series of models that clustered by the six districts in our sample. To do this, we implemented a form of the wild-cluster bootstrap procedure articulated in Cameron, Gelbach, and Miller (2008) for calculating robust standard errors when the number of clusters is small. Specifically, we used the Stata cgmwildboot.ado written by Judson Caskey at University

of California, Los Angeles (UCLA), Anderson. We are grateful for the thoughtful correspondence provided by Judson Caskey, Doug Miller, and Jonah Gelbach on this matter. The results of this approach, while yielding slightly larger standard errors in some cases and slightly smaller errors on other cases, did not change our overall conclusions. Moreover, McKinnon and Webb (2016), Webb (2014), and Cameron and Miller (2015) all caution against using the wild-cluster bootstrap procedure when the number of clusters is either very small or unbalanced, both of which are true in our case.

4. Classrooms with no variation in student race will not contribute to our estimates because there is no variation. Black teachers are most likely to teach a class that consists of 100% Black students (17%), followed by Hispanic teachers (13%). Less than 1% of White teachers (0.63%) teach a class that consists of 100% White students.

5. To test for differential sorting, we conduct a sorting test similar to Fairlie, Hoffmann, and Oreopoulos (2014) and Gershenson, Holt, and Papageorge (2016). We aggregate the data so that each regression contains just two observations per teacher, with the dependent variables being the White and non-White mean student characteristics, or the male and female mean student characteristics. We then run a series of regressions that includes indicators for student and teacher types (e.g., non-White teacher, non-White student) and the interaction of a non-White teacher and non-White student. The outcomes tested include prior math and English language arts (ELA) test scores, male student (in models examining racial sorting), minority students (in models examining gender sorting), gifted status, and special education status. The coefficient estimates of interest are those associated with the interaction term, all of which are insignificant at $p < .05$ in both ordinary least square (OLS) models and models in which we add grade and year fixed effects, alleviating concerns that our main results are potentially biased by student sorting into classrooms.

6. It should be noted that we do not apply any multiple-comparison corrections to our results. If we had, the degree to which we fail to reject the null hypothesis would be increased, as there are likely some spurious findings given the number of outcomes we examine.

7. To test the strength of the random assignment, we also ran models with this subsample that included no student controls, which made little difference to our findings. Furthermore, in the spirit of a Chow test, where we formally tested whether the OTHER estimate in the randomly assigned sample is significantly different from that in the nonrandomly assigned sample, we did not observe any indication that the results differed significantly.

8. We have also estimated models that combine related outcomes into three overarching factors: Press (Challenge, Control), Support (Confer, Captivate, Clarify, Consolidate), and Positive (Care, Happy, Effort), and the results for these models are consistent with what has been presented thus far, in that all three factors are consistently negative and significant.

9. A noteworthy example of the potential problem with these restricted models can be seen in the case of Black females. Column 12 presents results for Black female students. Other race, other sex (OROS) students (White males in this case) in the same classroom as Black female students rate Black female teachers −0.24 *SD* lower. However, we can also observe Black female students rating White male teachers in column 3 (the reverse of column 12), Black female students rating White female teachers in column 5 (the reverse of column 11), and Black female students rating Black male teachers in column 7 (the reverse of column 10). In all cases (columns 3, 5, and 7), the estimates are positive compared with the reference group (matches). This suggests, as do the descriptive characteristics in Table 2, that Black female students may tend to assign higher ratings on average. Generally speaking, there are very few instances in this set of results that raise concern, as there are very few positive coefficients in Table 7 that are statistically significant. Nonetheless, it is important to interpret these results with caution.

References

Adair, A. V. (1984). *Desegregation: The illusion of Black progress*. Lanham, MD: University Press of America.

Almlund, M., Duckworth, A. L., Heckman, J. L., & Kautz, T. D. (2011, February). *Personality psychology and economics* (NBER Working Paper No. 16822). Retrieved from http://www.nber.org/papers/w16822

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Beady, C. H., Jr., & Hansell, S. (1981). Teacher race and expectations for student achievement. *American Educational Research Journal*, *18*, 191–206.

The Bill & Melinda Gates Foundation. (2010). *Learning about teaching: Initial findings from the measures of effective teaching project*. Seattle, WA: Author. Retrieved from http://eric.ed.gov/?id=ED528382

Blazar, D., & Kraft, M. A. (2015). *Teacher and teaching effects on students' academic behaviors and mindsets* (Mathematica working paper). Retrieved from http://scholar.harvard.edu/files/david_blazar/files/blazar_and_kraft_teacher_and_teaching_effects_on_students_academic_behaviors_and_mindsets_mathematica_working_paper_41.pdf?m=1449506784

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

Boser, U. (2011). *Teacher diversity matters: A state-by-state analysis of teachers of color*. Washington, DC: Center for American Progress.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, *90*, 414–427.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, *50*, 317–372.

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

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*, 2633–2679.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources*, *41*, 778–820.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2009). The academic achievement gap in grades 3 to 8. *The Review of Economics and Statistics*, *91*, 398–419.

Cohen, G., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: Asocial-psychological intervention. *Science*, *313*, 1307–1310.

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

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

Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources*, *42*, 528–554.

Dee, T. S. (2015). Social identity and achievement gaps: Evidence from an affirmation intervention. *Journal of Research on Educational Effectiveness*, *8*, 149–168.

Downey, D. B., & Pribesh, S. (2004). When race matters: Teachers' evaluations of students' classroom behavior. *Sociology of Education*, *77*, 267–282.

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*, 237–251.

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.

Ehrenberg, R. G., Goldhaber, D., & Brewer, D. J. (1995). Do teachers' race, gender, and ethnicity matter? Evidence from the National Educational Longitudinal Study of 1988. *Industrial and Labor Relations Review*, *48*, 547–561.

Evans, M. O. (1992). An estimate of race and gender role-model effects in teaching high school. *The Journal of Economic Education*, *23*, 209–217.

Every Student Succeeds Act (ESSA) of 2015, Pub.L. 114-95 (2015).

Fairlie, R. W., Hoffmann, F., & Oreopoulos, P. (2014). A community college instructor like me: Race and ethnicity interactions in the classroom. *American Economic Review*, *104*, 2567–2591.

Ferguson, R. F. (1998). Teachers' perceptions and expectations and the Black-White test score gap. In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap* (pp. 217–317). Washington, DC: The Brookings Institution Press.

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

Ferguson, R. F. (n.d.). *The tripod project framework*. Retrieved from http://tripoded.com/survey-modules/

Ferguson, R. F., & Danielson, C. (2014). How framework for teaching and Tripod 7Cs evidence distinguish key components of effective teaching. In T. Kane, K. Kerr, & R. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the measures of effective teaching project* (pp. 98–143). San Francisco, CA: Jossey-Bass.

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

Fox, L. (2015). Seeing potential: The effects of student–teacher demographic congruence on teacher expectations and recommendations. *AERA Open*. Retrieved from http://ero.sagepub.com/content/2/1/2332858415623758.abstract

Fryer, R. G., Jr., & Torelli, P. (2010). An empirical analysis of "acting white." *Journal of Public Economics*, *94*, 380–396.

Gay, G. (2010). *Culturally responsive teaching: Theory, research, and practice*. New York, NY: Teachers College Press.

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

Gershenson, S., Hart, C. M. D., Lindsay, C. A., & Papageorge, N. W. (2017). *The long run effects of same-race teachers* (IZA Discussion Paper No. 10630). Retrieved from http://ftp.iza.org/dp10630.pdf

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.

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

Goldhaber, D., Theobald, R., & Tien, C. (2015). *The theoretical and empirical arguments for diversifying the teacher workforce: A review of the evidence* (CEDR Policy Brief No. 2015-920159). Retrieved from http://www.cedr.us/papers/working/CEDR%20WP%202015-9.pdf

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

Gregory, A., Skiba, R. J., & Noguera, P. A. (2010). The achievement gap and the discipline gap: Two sides of the same coin? *Educational Researcher*, *39*, 59–68.

Grissom, J. A., Kern, E. C., & Rodriguez, L. A. (2015). The "representative bureaucracy" in education: Educator workforce diversity, policy outputs, and outcomes for disadvantaged students. *Educational Researcher*, *44*, 185–192.

Grissom, J. A., Loeb, S., & Doss, C. (2015). The multiple dimensions of teacher quality: Does value-added capture teachers' nonachievement contributions to their schools? In J. A. Grissom & P. Youngs (Eds.), *Improving teacher evaluation systems: Making the most of multiple measures* (pp. 37–50). New York, NY: Teachers College Press.

Grissom, J. A., & Redding, C. (2016). Discretion and disproportionality: Explaining the underrepresentation of high-achieving students of color in gifted programs. *AERA Open*, *2*(1), 1–25.

Hanushek, E. A. (2011). Valuing teachers: How much is a good teacher worth? *Education Next*, *11*(3), 40–45.

Hanushek, E. A., Ingram, G. K., & Kenyon, D. A. (2014). Is location fate? Distributional aspects of schooling. In G. K. Ingram & D. A. Kenyon (Eds.), *Education, land, and location* (pp. 25–61). Cambridge, MA: Lincoln Institute of Land Policy. Retrieved from http://hanushek.stanford.edu/sites/default/files/publications/Hanushek%202014%20IsLocationFate%20%281%29.pdf

Heckman, J. J., & Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review*, *91*, 145–149.

Hess, F. M., & Leal, D. L. (1997). Minority teachers, minority students, and college matriculation: A new look at the role-modeling hypothesis. *Policy Studies Journal*, *25*, 235–248.

Holt, S. B., & Gershenson, S. (2015, December). *The impact of teacher demographic representation on student attendance and suspensions* (IZA Discussion Paper No. 9554). Retrieved from http://www.iza.org/en/webcontent/publications/papers/viewAbstract?dp_id=9554

Irvine, J. J. (2000). Afrocentric education: Critical questions for further considerations. In D. Pollard & C. Ajirotutu (Eds.), *African-centered schooling in theory and practice* (pp. 199–210). Westport, CT: Bergin & Garvey.

Jackson, C. K. (2012, December). *Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina* (NBER Working Paper No. 18624). Retrieved from http://www.nber.org/papers/w18624

Jackson, C. K. (2016, May). *What do test scores miss? The importance of teacher effects on non-test score outcomes* (NBER Working Paper No. 22226). Retrieved from http://www.nber.org/papers/w22226

Jennings, J. L., & DiPrete, T. A. (2010). Teacher effects on social and behavioral skills in early elementary school. *Sociology of Education*, *83*, 135–159.

Kane, T. J., & Cantrell, S. (2010). *Learning about teaching: Initial findings from the measures of effective teaching project*. Seattle, WA: The Bill & Melinda Gates Foundation.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment* (MET Project Research Paper). Seattle, WA: The Bill & Melinda Gates Foundation. Retrieved from http://files.eric.ed.gov/fulltext/ED540959.pdf

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

King, S. H. (1993). The limited presence of African-American teachers. *Review of Educational Research*, *63*, 115–149.

Koedel, C., Parsons, E., Podgursky, M., & Ehlert, M. (2015). Teacher preparation programs and teacher quality: Are there real differences across programs? *Education Finance and Policy*, *10*, 508–534.

Kraft, M. A., & Grace, S. (2016). *Teaching for tomorrow's economy? Teacher effects on complex cognitive skills and social-emotional competencies* (Working paper). Retrieved from http://scholar.harvard.edu/mkraft/publications/teaching-tomorrows-economy-teacher-effects-complex-cognitive-skills-and-socio

Ladson-Billings, G. (1992). Liberatory consequences of literacy: A case of culturally relevant instruction for African American students. *Journal of Negro Education*, *61*, 378–391.

Ladson-Billings, G. (1994). *The dreamkeepers: Successful teachers of African American students*. San Francisco, CA: Jossey-Bass.

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

Lindsay, C. A., & Hart, C. M. D. (2017). Teacher race and school discipline. *Education Next*, *17*(1), 72–78.

McCarthy, J. D., & Hoge, D. R. (1987). The social construction of school punishment: Racial disadvantage out of universalistic process. *Social Forces*, *65*, 1101–1120.

McKinnon, J. G., & Webb, M. D. (2016). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics*, *32*, 233–254.

Merrell, K. W., & Gueldner, B. A. (2010). *Social and emotional learning in the classroom: Promoting mental health and academic success*. New York, NY: Guilford Press.

Nixon, L. A., & Robinson, M. D. (1999). The educational attainment of young women: Role model effects of female high school faculty. *Demography*, *36*, 185–194.

Ouazad, A. (2014). Assessed by a teacher like me: Race and teacher assessments. *Education Finance & Policy*, *9*, 334–372.

Papageorge, N. W., Gershenson, S., & Kang, K. (2016, August). *Teacher expectations matter* (IZA Discussion Paper No. 10165). Retrieved from http://ftp.iza.org/dp10165.pdf

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

Pitts, D. (2007). Representative bureaucracy, ethnicity, and public schools. *Administration and Society*, *39*, 497–526.

Rampey, B. D., Dion, G. S., & Donahue, P. L. (2009). *NAEP 2008 tends in academic progress* (Report No. NCES 2009–479). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Robinson, J. P., & Lubienski, S. T. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher ratings. *American Educational Research Journal*, *48*, 268–302.

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, *3*, 16–20.

Ruzek, E. A., Domina, T., Conley, A. M., Duncan, G. J., & Karabenick, S. A. (2015). Using value-added models to measure teacher effects on students' motivation and achievement. *The Journal of Early Adolescence*, *35*, 852–882.

Snyder, T. (2016). *Mobile digest of education statistics, 2014* (NCES 2016-011). Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613–629.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*, 797–811.

Stewart, J., Meier, K. J., & England, R. E. (1989). In quest of role models: Change in Black teacher representation in urban school districts, 1968-1986. *Journal of Negro Education*, *58*, 140-152.

Tough, P. (2012). *How children succeed: Grit, curiosity, and the hidden power of character*. Boston, MA: Houghton Mifflin Harcourt.

U.S. Department of Education. (2015). *The condition of education 2015*. Washington, DC: U.S. Government Printing Office.

Villegas, A. M., Strom, K., & Lucas, T. (2012). Closing the racial/ ethnic gap between students of color and their teachers: An elusive goal. *Equity and Excellence in Education*, *45*, 283–301.

Wallace, L. W., Kelcey, B., & Ruzek, E. (2016). What can student perception surveys tell us about teaching? Empirically testing the underlying structure of the Tripod student perception survey. *American Education Research Journal*, *53*, 1834–1868.

Webb, M. D. (2014). *Reworking wild bootstrap based inference for clustered errors* (Queens Economics Department Working Paper No. 1315). Retrieved from http://qed.econ.queensu.ca/working_papers/papers/qed_wp_1315.pdf

Weinstein, C. S., Tomlinson-Clarke, S., & Curran, M. (2004). Toward a conception of culturally responsive classroom management. *Journal of Teacher Education*, *55*, 25–38.

West, M. R., Kraft, M. A., Finn, A. S., Duckworth, A. L., Gabrieli, C. F. O., & Gabrieli, J. D. E. (2016). Promise and paradox: Measuring students' non-cognitive skills and the impact of schooling. *Educational Evaluation and Policy Analysis*, *38*, 148–170.

Winters, M. A. (2011). *Teachers matter: Rethinking how public schools identify, reward, and retain great educators*. Summit, PA: Rowman & Littlefield.

Winters, M. A., Haight, R. C., Swaim, T. T., & Pickering, K. A. (2013). The effect of same-gender teacher assignment on student achievement in the elementary and secondary grades: Evidence from panel data. *Economics of Education Review*, *34*, 69–75.

Zirkel, S. (2002). Is there a place for me? Role models and academic identity among White students and students of color. *Teachers College Record*, *104*, 357–376.

Authors

ANNA J. EGALITE is an assistant professor in the Department of Educational Leadership, Policy, and Human Development at North Carolina State University. Her research focuses on the evaluation of education policies and programs intended to close racial and economic achievement gaps. Her studies have examined school choice policy, school size, the influence of family background on intergenerational economic mobility, and the diversification of the teacher labor force.

BRIAN KISIDA is an assistant research professor in the Department of Economics and the Truman School of Public Affairs at the University of Missouri. His research focuses on identifying effective educational options and experiences for at-risk students that can close the achievement gap, the experience gap, and the attainment gap. He has studied the educational benefits of cultural institutions, school–community partnerships, arts education, teacher diversity, and urban school choice.

Manuscript received June 7, 2016
First revision received February 9, 2017
Second revision received May 5, 2017
Accepted May 8, 2017