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

Anna J. Egalite
Department of Educational Leadership,
Policy, and Human Development
North Carolina State University
anna_egalite@ncsu.edu

Brian Kisida
Department of Economics &
Truman School of Public Affairs
University of Missouri
kisidab@missouri.edu

Abstract:
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.

.

Working Paper: Please do not cite without authors' permission.

Acknowledgements: We wish to thank Susan Dynarski, Raegan Miller, Seth Gershenson, and participants at the 2016 meeting of the Association for Education Finance and Policy for helpful feedback on earlier drafts of this paper.

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

Abstract:

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.

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

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, 2005; 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 congruence is related to gains in student achievement (e.g., Dee, 2004), increased perceptions of student engagement and performance (Dee, 2007; Ouazad 2014), higher expectations for educational attainment (Gershenson, Holt, & Papageorge, 2016), and reductions in absences and suspensions (Holt & Gershenson, 2015). 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, Theobald,

& Tien, 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), 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 non-cognitive outcomes (Heckman & Rubinstein, 2001), social emotional skills (Merrell & Gueldner, 2010) and academic behaviors and mindsets (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 & Gershenson's (2015) examination of the effect of student-teacher demographic mismatch on student attendance and suspensions (Holt & Gershenson, 2015), these 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 that are directly tied to their classroom teachers. While earlier work (e.g., Dee, 2007; Gershenson et al., 2016; Ouazad, 2014) investigates match effects on teachers'

perceptions of students, this paper 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 student 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 4-8 (Ferguson, 2008). The TRIPOD measures include scales indicating if a student feels cared for by his/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, Kelcey & Ruzek, 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 (Kraft & Blazar, 2016).

As such, these classroom-specific survey measures offer a unique window into some of the more specific ways in which a students' 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, 2015). 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 sub-sample of students in the MET project were randomly assigned to teachers, we are able to provide some additional assurance that our strategy addresses the potential bias that could result from the non-random assignment of students to teachers.

Our results suggest there are important benefits for students' academic perceptions and attitudes 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 to non-matched 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 in order to better serve students of color (Goldhaber, 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 et al., 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 & Clewell, 1998; Villegas & Lucas, 2004). Exposure to a successful mentor in this manner 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). Further, 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, Meier, & England, 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, Holt, & Papageorge, 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 absolute level of quality and expectation 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." Further, if teachers are familiar with students' cultural backgrounds, they might be less likely to succumb to intentional or unintentional biases stemming from negative stereotypes that alter the ways that teachers interact with students who have different demographic characteristics (e.g., Ferguson, 1998), especially if they hold stereotypes related to perceived academic ability (Rosenthal & Jacobson, 1968). Finally, a strong interpersonal connection between student and teacher stemming from a shared cultural understanding might reduce the likelihood of suspension, expulsion, or other extreme disciplinary responses to student misbehavior, which has previously been shown to be susceptible to biases along racial or gender lines (Downey & Pribesh, 2004; Gregory, Skiba, & Noguera, 2010; Lindsay & Hart, 2016; 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 of own-race teacher matches 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 et al. (2007) and Goldhaber and Hansen (2010), both use longitudinal data from North Carolina to document student achievement effects from racially congruent teachers of 0.02 to 0.03 standard deviations. Using the same data, Goldhaber and Hansen find that black students with black teachers experienced the largest gains, at roughly 0.04 standard deviations. Similarly, employing a student fixed-effects analysis with eight years of data, Authors find some evidence of student-teacher matching effects in the range of 0.01 to 0.04 standard deviations, 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 National Education Longitudinal Study of 1988 (NELS:88), Dee finds that assignment to an opposite-gender English teacher for one year reduces student achievement by 0.05 standard deviations. Winters, Haight, Swaim, and Pickering

(2013) analyze an administrative panel dataset from Florida, also finding no significant relationship from student-teacher gender interaction. Rather, they conclude that both male and female students benefit from being assigned to a female teacher.

Effects on Non-Tested Academic Outcomes

Research has also examined student-teacher demographic congruence on subjective or "non-tested" measures. Using the NELS:88 data, which includes 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 standard deviations. 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 (2007), Gershenson, Holt and Papageorge (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 nonblack teachers are 12 percentage points more likely to expect black students will only complete a high school diploma or less. Gershenson and colleagues find no effects from student-teacher gender interactions. 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 Non-Tested Academic Outcomes and Teacher Effectiveness

We also draw on the broader literature on 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, Freidman, & Rockoff, 2014). Though 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 non-tested 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 mindset, effort in class, and grit. Yet, similar to Blazar and Kraft (2016), 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* is a particularly telling manifestation of the growing reluctance to rely on narrowly-defined measures of teacher and school effectiveness (ESSA, 2015), though some researchers are skeptical that newly adopted approaches have been thoroughly vetted (Duckworth & Yeager, 2015). Though the literature to date that attempts to demonstrate teachers' impact on students' non-cognitive 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 et al., 2011; Heckman & Rubenstein, 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 current trends aimed at better understanding teacher quality.

Data and Measures

The MET Project

Our data are drawn from the Measures of Effective Teaching (MET) project, a project of the Bill and Melinda Gates Foundation which tracked approximately 3,000 teachers in six school districts across the United States over the 2009-10 and 2010-11 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 two 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. Though 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 tend to have fewer years' 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 student rosters to MET project teachers. In order 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 to "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). Though 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 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 paper, 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 identified all students in grades four through eight—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, nearly 11,000 students, or 12 percent of the sample each year are duplicates[1].

Table 1 presents a descriptive overview of the sample. Twenty-eight percent of students are white, 38 percent are Black, and 34 percent 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 percent) qualify for the federal free and reduced price lunch program. Thirteen percent of students are classified as English Language Learners, 10 percent are gifted, and 9 percent 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 percent of teachers are male, 57 percent are white, 37 percent are Black, and 6 percent are Hispanic. (Appendix Table A1 provides these student and teacher summary statistics broken out for each of the six districts under study).

≪ TABLE ONE ABOUT HERE ≫

Table 1 also presents summary statistics for the ten dependent variables, whose values range from one to five. The full list of items included in each scale and the associated Cronbach's Alphas are listed in Appendix B. Further, Appendix Table C1 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 ten 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 percent 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.

≪ TABLE TWO ABOUT HERE ≫

Our independent variables include "Same Race" and "Same Sex." We also include other possible combinations in subsequent models (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 percent, and 9 percent for Hispanic students. All of these student-level percentages exceed the share of teachers' representation in our sample (57 percent of teachers are white, 37 percent are black, and 6 percent are Hispanic), illustrating the dramatic non-random sorting of students and teachers across classrooms and schools. The aggregate statistic for assignment to a same-sex teacher (50 percent) 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.

≪ TABLE THREE ABOUT HERE ≫

Identification Strategy

The primary identification strategy builds upon that presented by Dee (2007) and Gershenson, Holt, and Papageorge (2016), which leverages contemporaneous subjective evaluations of students 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 academic perceptions and attitudes (*APA*) for student, *i*, in classroom, *j*, is measured as:

$$APA_{ij} = \beta_0 + \beta_1 \boldsymbol{Other}_i + \beta_2 \boldsymbol{X}_i + \alpha_j + \epsilon_{ij} \qquad (1)$$

where ***Other*** is a vector of variables that measure demographic mismatch between teacher and student. Following 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, Holt and Papageorge (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 math and reading test scores from 2009, prior to the start of the MET data collection period). These variables capture systematic differences in student's 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 free or reduced price lunch[2] and English language learners—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 teacher 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., 2015). 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. Additionally, 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. Additionally, 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].

Lastly, 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, 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. Though we note this caveat, we find no evidence to suggest that differential sorting is likely to pose a significant problem.

Results

For our initial examination of the effect of teacher/ student demographic match on students' academic perceptions and attitudes, 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 of the scales examined, with significant effect sizes ranging from -0.02 to -0.06 standard deviations (hereafter, 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 same gender of the teacher. Similarly, we observe statistically significant negative coefficients on the *Other Race* variable for Care, Happy, Clarify, Control, and Consolidate that range from -0.02 to -0.04 SD. On average, students whose teacher is the same race/ethnicity report having more favorable perceptions of their teacher than students in the same classroom who do share their teacher's race/ethnicity[5].

≪ TABLE FOUR ABOUT HERE ≫

Columns 3 and 4 add controls for math and ELA test scores in 2009, which is the year prior to the start of the MET project. This restriction reduces the sample size due to some missing prior test scores in the data. Using this restricted sample, the results are nearly identical, with significant effects sizes ranging from -0.03 to -0.07 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 non-random sorting of students to class sections. Thus, we sacrifice a sample size of approximately 70,000 observations for one closer to 12,800. Nonetheless, the randomly assigned sample 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[6].

In summary, across all three specifications, the largest and most consistently significant effects are observed for Care, Captivate, Confer, Challenge, and Consolidate. In general, the effects are largest and more consistent among gender matches[7].

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 Other Sex, *Other Race and Same Sex*, and *Other Race and Other 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 SROS range from -0.04 to -0.08 SD, whereas the coefficients on ORSS range from -0.02 to -0.05 SD. Interestingly, 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 (OROS). The statistically significant coefficients range from -0.04 to -0.13 SD. Finally, as we observed before, the results from the randomly-assigned sample are well-aligned with the results that rely on exclusively on classroom fixed-effects.

≪ TABLE FIVE ABOUT HERE ≫

Elementary and Middle School Subgroups

It is also instructive to break out the impacts on academic perceptions and attitudes 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, while the results seem to be generally larger for middle school students when race matching is included (ORSS and OROS). Examples include Care—where the SROS coefficient is -0.06 SD for elementary students and -0.07 SD for middle school students— and Clarify— where the SROS coefficient is -0.08 SD for elementary students and -0.06 SD for middle school students. The estimates for Happy are also consistent across elementary and middle school students in the SROS columns. Other variables such as Captivate

have similar point estimates (e.g., the SROS coefficient is -0.05 SD for the elementary sample and -0.07 SD for the middle school sample) but are statistically significant at conventional levels in the middle school sample only. This disparity may be explained by differences in sample size and the resulting impact on statistical power, given that the middle school sample is approximately twice as large as the elementary sample. The largest divergence from this pattern is that middle school students matched on gender are more likely to report thinking about college more because of their teacher. This pattern is similarly reflected among middle school students when race matching is included. 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.

≪ TABLE SIX ABOUT HERE ≫

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 -0.13 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's 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 mismatch.

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 nine out of ten outcomes, with coefficients ranging from -0.05 to -

0.10 SD. Thus, while both age groups appear to be impacted by assignment to a demographically-similar teacher, it is students in grades six through eight 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 male teachers, and black female teachers. In order to provide the most reliable estimates, all estimates in Table 7 are based on the sample of teachers 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 mismatches; thus, we exclude Hispanic teachers and students from the table for ease of presentation. For example, in the case of columns 1 through 3, which are restricted to white male teachers, all effects are relative to a white male student. 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. 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. Additionally, 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, 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[8].

≪ TABLE SEVEN ABOUT HERE ≫

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, which show up across all types of teachers in 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 large negative "reverse" effects in column 9.

In terms of particular outcome variables, few notable trends emerge. Most results for Care, Confer, Effort, Challenge, and Consolidate, follow a general trend of being significant when matched students are compared to same race students of different gender and same race in the same classroom, with effects that increase when matched students are compared to 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 ten academic perceptions and attitudes. 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 to other 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' academic perceptions and attitudes and may shed light on the specific ways in which students are affected by the wide demographic divide in American public education. The largest effects tend to be concentrated when students share both gender and racial characteristics with their teachers, compared to 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 towards a majority-minority student population. Between 1992 and 2012, the proportion of students who were white dropped from 67 to 51 percent whereas the proportion of students who were Hispanic rose from 12 to 24 percent and the proportion of students who were

black held steady at roughly 16 percent (Snyder, 2016). During this same time period, however, the teacher workforce did not undergo a similar demographic transformation, maintaining a white, female majority.

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 to 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 scale and Effort scale. These measures capture student reports about their teacher pushing them to work hard, accepting nothing less than their full effort, and whether they feel they are doing their best quality work. These outcomes are closely related to prior theory that teachers may hold particularly "high 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, Holt, & Papageorge, 2016; Ouazad, 2014).

What we find particularly compelling is the evidence we uncover in support of the theory of cultural understanding, which makes the case 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 to 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. We also see large effects for black male students in the Care, Confer, and Control scales. 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 don't understand something, and 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/she really cares about me" align nicely with this theory.

Our results and additional recent research (e.g., Holt & Gershenson, 2015) research suggest that studies focused on achievement effects may have only observed the tip of the iceberg with regards to the benefits of demographically similar teachers. These findings 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 potential reasons for 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 students' 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 [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)] are urban districts and these same results may not hold in other locations. Second, the measures relied upon in this study 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. 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. Only then can the full implications of the teacher gender gap and the teacher diversity gap be fully considered.

References

Almlund, M., Duckworth, A. L., Heckman, J. L., & Kautz, T. D. (2011). Personality psychology and economics. National Bureau of Economic Research working paper 16822, Retrieved June 26, 2014 (http://www.nber.org/papers/w16822).

Antecol, H., Eren, O., & Ozbeklik, S. (2015). The effect of teacher gender on student achievement in primary school. *Journal of Labor Economics*, 33(1), 63-89.

Bates, L. A., & Glick, J. E. (2013). Does it matter if teachers and schools match the student? Racial and ethnic disparities in problem behaviors. *Social Science Research*, *42*(5), 1180–1190.

Beady, C. H., Jr., Hansell, S. (1981). Teacher race and expectations for student achievement. *American Educational Research Journal,* 18(2): 191-206.

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.

Boser, U., Wilhelm, M., & Hanna, R. (2014). *The power of the Pygmalion effect: Teachers expectations strongly predict college completion*. Washington, DC: Center for American Progress.

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. *Review of Economics and Statistics*, *90*(3), 414–427.

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

Chetty, R., Friedman, J. N., & Rockoff, J.E., 2013. Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9), 2633-79.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2009). The academic achievement gap in grades 3 to 8. *The Review of Economics and Statistics*, *91*(2), 398–419.

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

Cornwell, C., Mustard, D. B., & Van Parys, J. (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources*, 48(1), 236-264.

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. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources*, 42(3), 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*(4), 267–282.

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*(4), 237–251.

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(3), 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(3), 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(8), 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(4), 460–507.

Ferguson, R. F. (2008). The Tripod Project Framework. Retrieved from http://www.achievementseminars.com/seminar_series_2008_2009/readings/ferguson_tripod_project.pdf.

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: Jossey-Bass.

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

Garrett, R., & Steinberg, Matthew P. (2015). Examining teacher effectiveness using classroom observation scores: Evidence from the randomization of teachers to students. *Educational Evaluation and Policy Analysis*, 37(2), 224-242.

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

Gershenson, S. (2016). Linking teacher quality, student attendance, and student achievement. *Education Finance and Policy*, *11*(2), 125–149.

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

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(1), 59–68.

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

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*(3), 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.

Hanushek, E. A., Ingram, G. K., & Kenyon, D. A. (2014). Is location fate? Distributional aspects of schooling. In *Education, Land, and Location* (pp. 25–61). Cambridge, MA: Lincoln Institute of Land Policy.

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

Heckman, J., J., & Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review*, *91*(2), 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(2), 235–248.

Holt, S. B., & Gershenson, S. (2015). 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 and C. Ajirotutu (Eds.), *African-centered schooling in theory and practice* (pp. 199-210). Westport, CT: Bergin & Garvey.

Jackson, C. K. (2016). 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

Jackson, C. K. (2012). 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.

Jennings, J. L., & DiPrete, T. A. (2010). Teacher effects on social and behavioral skills in early elementary school. *Sociology of Education, 83*(2), 135–159.

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

Kane, T. J., McCaffrey, D. F., Miller, T., 7 Staiger, D. O. (2013). Have we identified effective teachers? Validating measures of effective teaching using random assignment. *MET Project Research Paper*. Seattle, WA: 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: The Bill & Melinda Gates Foundation.

King, S.H. (1993). The limited presence of African-American teachers, *Review of Educational Research.* 63(2): 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(4), 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. (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(3), 465–491.

Lavy, V. (2008). Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment. *Journal of Public Economics*, 92(10), 2083-2105.

Lindsay, C. & Hart, C. (2017). Teacher race and school discipline. *Education Next*, 17(1).

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(2), 185–194.

Ouazad, A. (2014). Assessed by a teacher like me: Race and teacher assessments. *Education Finance & Policy*, 9(3), 334-372.

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.

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(2), 268–302.

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, 3(1), 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*(5), 852–882.

Snyder, T. (2016). *Mobile Digest of Education Statistics, 2014* (NCES 2016-011). Washington D.C.: National Center for Education Statistics, U.S. Department of Education.

The Bill & Melinda Gates Foundation. (2010). Learning about teaching: Initial findings from the Measures of Effective Teaching Project. Seattle, WA: The Bill and Melinda Gates Foundation. Retrieved http://eric.ed.gov/?id=ED528382

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

U.S. Department of Education (2015). *The Condition of Education 2015*, Washington D.C.: U.S. Government Printing Office.

U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service. (2016). *The state of racial diversity in the educator workforce*. Washington, D.C. Retrieved from

http://www2.ed.gov/rschstat/eval/highered/racial-diversity/state-racial-diversity-workforce.pdf

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(6), 1834-1868.

Weinstein, C. S., Tomlinson-Clarke, S., & Curran, M. (2004). Toward a conception of culturally responsive classroom management. *Journal of Teacher Education*, 55(1), 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(1): 148-170.

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

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*(2), 357–376.

Table 1.
Descriptive Statistics of the Analytic Sample

Scale	Mean	SD	Min	Max	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
Masters/ 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

Notes: n = 82,409 unique students, 1,909 teachers, 231 schools

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

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

Notes: n = 82,409 students, 1,909 teachers, 231 schools

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*	-.07***	-.05*	-.08***
	(.01)	(.01)	(.02)	(.01)	(.03)	(.03)
	n = 69,852		n = 55,907		n = 12,796	
Captivate	-.01	-.06***	-.02	-.06***	-.01	-.06***
	(.01)	(.01)	(.02)	(.01)	(.03)	(.02)
	n = 70,054		n = 56,100		n =12,825	
Happy	-.02	-.04***	-.03**	-.04***	-.06*	-.04*
	(.01)	(.01)	(.02)	(.01)	(.04)	(.03)
	n = 69,507		n = 55,714		n = 12,754	
Confer	-.02	-.04***	-.02	-.04***	-.02	-.05**
	(.01)	(.01)	(.02)	(.01)	(.03)	(.02)
	n = 70,048		n = 56,102		n = 12,821	
Effort	-.03**	-.02*	-.03*	-.02*	-.03	-.04
	(.01)	(.01)	(.02)	(.01)	(.04)	(.03)
	n = 69,927		n = 55,985		n = 12,806	
College	-.02	-.06***	-.02	-.06***	n/a	n/a
	(.02)	(.01)	(.02)	(.01)		
	n = 46,236		n = 42,851			
Clarify	-.03**	-.04***	-.03**	-.05***	-.02	-.03
	(.01)	(.01)	(.02)	(.01)	(.03)	(.02)
	n = 69,941		n = 55,993		n = 12,804	
Control	-.04***	-.01	-.04***	-.01	-.02	-.03
	(.01)	(.01)	(.01)	(.01)	(.03)	(.02)
	n = 69,912		n = 55,990		n = 12,807	
Challenge	-.01	-.02**	-.02	-.04***	.00	-.06**
	(.01)	(.01)	(.02)	(.01)	(.03)	(.02)
	n = 70,001		n = 56,058		n = 12,822	
Consolidate	-.02*	-.03***	-.03**	-.03***	-.03	-.05**
	(.01)	(.01)	(.02)	(.01)	(.04)	(.03)`
	n = 69,488		n = 55,655		n = 12,763	
Prior Test Scores	No		Yes		Yes	

Notes: Models include controls for student gender, student race, the interactions of student gender by student race, FRL, ELL Sp.Ed., gifted, 2009 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 teacher level; *** p<0.01, ** p<0.05, and * p<0.1.

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*	-.10***	-.06*	-.03	-.13***
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.04)	(.04)	(.04)
Captivate	-.08***	-.02	-.07***	-.08***	-.04*	-.08***	-.06*	-.01	-.06
	(.01)	(.02)	(.02)	(.02)	(.02)	(.02)	(.03)	(.04)	(.04)
Happy	-.06***	-.03*	-.07***	-.06***	-.05***	-.08***	-.03	-.05	-.10**
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.04)	(.04)	(.04)
Confer	-.04***	-.02*	-.06***	-.05***	-.02	-.06***	-.06*	-.04	-.07*
	(.01)	(.01)	(.02)	(.02)	(.02)	(.02)	(.03)	(.04)	(.04)
Effort	-.03*	-.04**	-.05***	-.03	-.03*	-.05**	-.04	-.03	-.07
	(.01)	(.02)	(.02)	(.02)	(.02)	(.02)	(.04)	(.04)	(.05)
College	-.07***	-.03	-.08***	-.07***	-.03	-.08***	n/a	n/a	n/a
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)			
Clarify	-.05***	-.04**	-.07***	-.06***	-.04**	-.08***	-.03	-.03	-.06
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.03)	(.04)	(.04)
Control	-.02	-.05***	-.05***	-.02	-.05***	-.05***	-.02	-.01	-.05
	(.01)	(.01)	(.02)	(.01)	(.02)	(.02)	(.03)	(.04)	(.04)
Challenge	-.04***	-.03*	-.04**	-.05***	-.04**	-.06***	-.04	-.02	-.05
	(.01)	(.02)	(.02)	(.02)	(.02)	(.02)	(.03)	(.04)	(.04)
Consolidate	-.04**	-.03*	-.06***	-.04**	-.03*	-.06***	-.02	-.00	-.08*
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.03)	(.04)	(.04)
Prior Test Scores		No			Yes			Yes	

Notes: SROS = Same Race, Other Sex; ORSS = Other Race, Same Sex; OROS = Other Race, Other Sex. Models include controls for student gender, student race, the interactions of student gender by student race, FRL, ELL Sp.Ed., gifted, 2009 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 teacher level. Specific sample sizes are identical to those reported in Table 4. In columns 1 -3, *n* ranges from 46,236 to 70,054. In columns 4-6, *n* ranges from 42,851 to 56,102. In columns 7 - 9, *n* ranges from 12,763 to 12,825; *** p<0.01, ** p<0.05, and * p<0.1.

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**	-.07***	.02	-.04*	-.07*	-.09***
	(.03)	(.02)	(.03)	(.02)	(.04)	(.02)
Captivate	-.05*	-.07***	-.03	-.04	-.06	-.08***
	(.03)	(.02)	(.03)	(.02)	(.04)	(.02)
Happy	-.06*	-.05**	-.02	-.05**	-.04	-.08***
	(.03)	(.02)	(.03)	(.02)	(.04)	(.02)
Confer	-.04	-.04**	-.04	-.02	-.10**	-.05**
	(.03)	(.02)	(.03)	(.02)	(.04)	(.02)
Effort	-.03	-.02	-.01	-.04*	-.02	-.06**
	(.03)	(.02)	(.03)	(.02)	(.04)	(.03)
College	-.03	-.09***	.04	-.06**	-.03	-.10***
	(.04)	(.02)	(.04)	(.03)	(.04)	(.03)
Clarify	-.08**	-.06**	-.06*	-.03	-.13**	-.06***
	(.03)	(.02)	(.03)	(.02)	(.04)	(.02)
Control	-.01	-.02	-.06**	-.04	-.07**	-.04
	(.03)	(.02)	(.03)	(.02)	(.03)	(.02)
Challenge	-.10***	-.05**	-.05	-.03	-.06	-.07***
	(.03)	(.02)	(.03)	(.02)	(.04)	(.02)
Consolidate	-.05	-.05*	.01	-.06**	-.07*	-.07***
	(.04)	(.02)	(.03)	(.02)	(.04)	(.02)

Note: Separate regressions were run for elementary and middle school grades. Columns 1, 3, 5 come from the elementary grades regression (n = 18,531); columns 2, 4, 6 come from the middle school grades regression (n = 37,623). Models include controls for student gender, student race, the interactions of student gender by student race, FRL, ELL Sp.Ed., gifted, 2009 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 teacher level; *** p<0.01, ** p<0.05, and * p<0.1.

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		
Scale	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
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Care	-.09	-.07	-.16*	-.07	-.02	-.13***	-.08	-.18	-.28**	-.09**	-.06	-.18**
	(.13)	(.10)	(.09)	(.04)	(.05)	(.05)	(.12)	(.13)	(.14)	(.04)	(.08)	(.08)
Captivate	-.10	.03	-06	-.11**	-.01	-.05	.06	-.06	-.11	-.04	-.12	-.18**
	(.12)	(.09)	(.09)	(.05)	(.05)	(.05)	(.10)	(.16)	(.16)	(.04)	(.10)	(.09)
Happy	-.06	-.05	-.11	-.14***	-.04	-.12**	.08	-.09	-.12	-.04	-.09	-.13
	(.11)	(.10)	(.10)	(.04)	(.05)	(.05)	(.12)	(.14)	(.15)	(.04)	(.09)	(.08)
Confer	-.07	-.06	-.05	-.07*	.01	-.05	-.07	-.25*	-.30**	-.14***	-.09	-.18**
	(.10)	(.10)	(.09)	(.04)	(.05)	(.05)	(.14)	(.14)	(.14)	(.04)	(.08)	(.08)
Effort	.13	-.14	.03	-.18***	-.01	-.23***	.03	-.20	-.08	-.15***	-.01	-.24***
	(.10)	(.10)	(.10)	(.04)	(.05)	(.05)	(.12)	(.21)	(.24)	(.04)	(.09)	(.09)
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	-.15	-.02	.04	-.01	.07	-.01	.14	-.00	.00	-.15***	-.17**	-.30***
	(.12)	(.11)	(.10)	(.05)	(.05)	(.05)	(.11)	(.12)	(.14)	(.04)	(.08)	(.07)
Control	-.04	.08	-.01	-.03	.01	.03	-.13	-.25**	-.26**	.05	-.04	-.06
	(.08)	(.10)	(.10)	(.04)	(.04)	(.04)	(.10)	(.11)	(.13)	(.04)	(.08)	(.07)
Challenge	-.03	.08	.05	-.06	.06	-.04	.09	-.00	-.07	-.13***	-.13	-.26***
	(.11)	(.11)	(.09)	(.05)	(.05)	(.05)	(.11)	(.16)	(.17)	(.04)	(.08)	(.08)
Consolidate	-.13	-.01	-.07	.02	.08	-.01	.05	-.21	-.25*	-.08**	-.18**	-.30***
	(.12)	(.10)	(.09)	(.04)	(.05)	(.05)	(.10)	(.15)	(.14)	(.04)	(.08)	(.08)

Notes: Models include controls for student gender, student race, FRL, ELL Sp.Ed., gifted, prior 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 teacher level; *** p<0.01, ** p<0.05, and * p<0.10. In columns 1-3, *n* ranges from 1,360 to 1,366. In columns 4-6, *n* ranges from 5,702 to 5,727. In columns 7-9, *n* ranges from 770 to 773. In columns 10-12, *n* ranges from 4,104 to

4,132; 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. *** p<0.01, ** p<0.05, and * p<0.1.

Appendix A

Appendix Table A1.
Sample Means of the Analytic Sample, by District

Scale	District 1	District 2	District 3	District 4	District 5	District 6	Grand Mean
Dependent Variables							
Care	3.67	4.18	3.89	3.70	3.79	3.47	3.71
Captivate	3.50	3.64	3.70	3.56	3.68	3.61	3.59
Happy	3.87	3.99	3.83	3.82	3.84	3.74	3.83
Confer	3.59	4.24	3.85	3.72	3.82	3.34	3.68
Effort	4.03	4.05	4.13	4.04	4.00	3.84	4.01
College	3.61	4.20	4.11	3.70	3.81	3.69	3.77
Clarify	3.94	4.24	4.12	3.96	4.07	3.83	3.99
Control	3.58	3.50	3.39	3.34	3.32	3.34	3.42
Challenge	4.03	4.23	4.20	4.16	4.16	4.06	4.11
Consolidate	3.66	3.97	3.86	3.70	3.93	3.81	3.78
Student Characteristics							
Male	0.49	0.50	0.50	0.50	0.50	0.51	0.50
White	0.53	0.17	0.06	0.50	0.18	0.03	0.28
Black	0.17	0.16	0.88	0.41	0.33	0.25	0.38
Hispanic	0.29	0.67	0.06	0.15	0.49	0.72	0.34
FRL	0.53	0.68	0.23	n/a	0.65	0.90	0.57
ELL	0.10	0.40	0.04	0.14	0.09	0.28	0.13
Gifted	0.08	0.09	0.04	0.17	0.10	0.10	0.10
Special Education	0.08	0.11	0.05	0.07	0.11	0.11	0.09
Grade	6.01	4.45	5.39	6.08	6.23	7.05	6.09
Teacher Characteristics							
Male	0.15	0.28	0.11	0.16	0.20	0.25	0.18
White	0.78	0.91	0.21	0.70	0.65	0.35	0.57
Black	0.13	0.03	0.79	0.27	0.27	0.57	0.37
Hispanic	0.08	0.06	0.00	0.02	0.08	0.09	0.06
Years of Experience	10.69	6.97	n/a	n/a	n/a	9.46	10.01
Masters/ Advanced Degree	0.16	0.53	0.75	0.36	n/a	0.28	0.36
Proportion of Overall Sample From This District	0.26	0.03	0.17	0.21	0.17	0.17	1.00

Notes: n = 82,409 students, 1,909 teachers, 231 schools. "n/a" means "not available." This is because some of the cooperating districts declined to share data on free /reduced price lunch status, teacher experience, or teacher's masters/advanced degree with the MET project. Teacher characteristics are constant within classroom so the omission of any teacher-level controls will not affect our estimates. Student FRL status is included as a control variable in later tables, however, so we follow Garrett and Steinberg (2015) by controlling for FRL with a revised variable that assigns a zero to missing values and including a missing value indicator variable.

Appendix B

Care Scale

Elementary Items (α = .84):
I like the way my teacher treats me when I need help.
My teacher is nice to me when I ask questions.
My teacher in this class makes me feel that he/she really cares about me.
If I am sad or angry, my teacher helps me feel better.
The teacher in this class encourages me to do my best
My teacher seems to know if something is bothering me.
My teacher gives us time to explain our ideas.

Secondary Items (α = .78):
My teacher seems to know if something is bothering me.
My teacher really tries to understand how students feel about things.
My teacher in this class makes me feel that s/he really cares about me.

Captivate Scale

Elementary Items (α = .69):
School work is interesting.
We have interesting homework
Homework helps me learn.
School work is not very enjoyable. (Do you agree?)*

Secondary Items (α = .83):
I like the ways we learn in this class.
My teacher makes lessons interesting.
My teacher makes learning enjoyable.
This class does not keep my attention--I get bored.*

Happy Scale

Elementary Items (α = .65):
This class is a happy place for me to be.
Being in this class makes me feel sad or angry.*

Secondary Items (α = .63):
This class is a happy place for me to be.
Being in this class makes me feel angry.*

Confer Scale

Elementary Items (α = .76):
When he/she is teaching us, my teacher asks us whether we understand
My teacher asks questions to be sure we are following along when he/she is teaching
My teacher checks to make sure we understand what he/she is teaching us.
My teacher tells us what we are learning and why.
My teacher wants us to share our thoughts

Students speak up and share their ideas about class work.
My teacher wants me to explain my answers -- why I think what I think

Secondary Items (α = .70):
Students speak up and share their ideas about class work.
My teacher gives us time to explain our ideas.
Students get to decide how activities are done in this class.
My teacher wants us to share our thoughts.

Effort Scale
Elementary (α = .53)
I have pushed myself hard to understand my lessons in this class.
I have done my best quality work in this class
When doing schoolwork for this class, I try to learn as much as I can and I don't worry about how long it takes.
In this class, I take it easy and do not try very hard to do my best.*
In this class, I stop trying when the work gets hard.*
I am happy with how well I have done in this class.
Overall, between homework, reading, and other class assignments, I worked hard in this class.

Secondary Items (α = .74):
I have pushed myself hard to completely understand my lessons in this class.
I have done my best quality work in this class all year long
When doing schoolwork for this class, I try to learn as much as I can and I don't worry about how long it takes.
In this class, I take it easy and do not try very hard to do my best.*
In this class, I stop trying when the work gets hard.*
Overall, between homework, reading, and other class assignments, I worked hard in this class.

College Scale
Elementary (α = .72) and Secondary Items (α = .78):
My teacher makes me want to go to college.
Because of my teacher, I think more about going to college.

Clarify Scale

Elementary Items (α = .78):

If you don't understand something, my teacher explains it another way
In this class, we learn to correct our mistakes.
My teacher explains difficult things clearly
My teacher has several good ways to explain each topic that we cover in this class
My teacher knows when the class understands, and when we do not.
My teacher explains things in very orderly ways.
I understand what I am supposed to be learning in this class
This class is neat -- everything has a place and things are easy to find.

Secondary Items (α = .79):

My teacher explains difficult things clearly.
When s/he is teaching us, my teacher thinks we understand even when we don't.*
My teacher has several good ways to explain each topic that we cover in this class
If you don't understand something, my teacher explains it another way
My teacher knows when the class understands, and when we do not.

Control Scale

Elementary Items (α = .60):

Our class stays busy and does not waste time.
My classmates behave the way my teacher wants them to.
Students behave so badly in this class that it slows down our learning*
Everybody knows what they should be doing and learning in this class.

Secondary Items (α = .84):

Student behavior in this class makes the teacher angry.*
My classmates behave the way my teacher wants them to.
I hate the way that students behave in this class.*
Student behavior in this class is under control.
Student behavior in this class is a problem.*
Students in this class treat the teacher with respect.
Our class stays busy and doesn't waste time

Challenge Scale

Elementary Items (α = .63):

My teacher pushes everybody to work hard.
In this class, my teacher accepts nothing less than our full effort.
My teacher pushes us to think hard about things we read.
In this class we have to think hard about the writing we do.

Secondary Items (α = .82):

In this class, we learn to correct our mistakes.
In this class, my teacher accepts nothing less than our full effort.
In this class, we learn a lot almost every day.
My teacher wants me to explain my answers -- why I think what I think.
My teacher doesn't let people give up when the work gets hard.
My teacher asks questions to be sure we are following along when s/he is teaching.
My teacher asks students to explain more about answers they give.

Consolidate Scale

Elementary Items (α = .52):

My teacher takes the time to summarize what we learn each day.
When my teacher marks my work, he/she writes on my papers to help me understand how to do better

Secondary Items (α = .79):

My teacher checks to make sure we understand what s/he is teaching us.
My teacher takes the time to summarize what we learn each day.
We get helpful comments to let us know what we did wrong on assignments.
The comments that I get on my work in this class help me understand how to improve.

* Items are reverse coded

Response Scale:

1: Totally Untrue
2: Mostly Untrue
3: Somewhat
4: Mostly
5: Totally True

Appendix C

Appendix Table C1.
Correlations among the Dependent Variables

	College	Effort	Confer	Happy	Captivate	Care	Clarify	Control	Challenge	Consolidate
College	1.00									
Effort	0.42	1.00								
Confer	0.50	0.41	1.00							
Happy	0.43	0.45	0.48	1.00						
Captivate	0.51	0.46	0.52	0.62	1.00					
Care	0.56	0.45	0.70	0.60	0.61	1.00				
Clarify	0.53	0.49	0.66	0.58	0.64	0.71	1.00			
Control	0.29	0.34	0.39	0.45	0.40	0.39	0.44	1.00		
Challenge	0.48	0.46	0.53	0.42	0.50	0.54	0.63	0.35	1.00	
Consolidate	0.48	0.39	0.55	0.48	0.58	0.60	0.65	0.34	0.58	1.00

Note: Data on ‘College’ come from 2009-10 only; all other correlations are calculated across two years of data (2009-10 and 2010-11), n = 82, 409

Endnotes

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 & Steinberg (2015) who use a missing data dummy variable when examining these same data to retain these observations. Our results when excluding this district our 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 (Kane, et al., 2013, Koedel et al., 2015, Garrett & Steinberg, 2015). However, our standard errors may be inflated if errors are also clustered within schools or districts. We test the former by directly estimating all specifications by clustering standard errors at the school level and our results are unaffected. To test for district clustering, we implement a form of the wild-cluster bootstrap procedure articulated in Cameron, Gelbach, & 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 U.C.L.A., Anderson. However, computing resources would not permit us to run our full model (which includes approximately 3,000 classroom dummy variables). As an alternative approach, we ran our models using school fixed effects and teacher-level clustering and compared these results to identical school-fixed effects models that implemented district-level clustering. The results of this comparison, while yielding slightly larger standard errors, did not change our overall conclusions. We are grateful for the thoughtful correspondence provided by Judson Caskey, Doug Miller, and Jonah Gelbach on this matter. On a related point, theoretically we could also have clustered at the student level because there are certainly students that appear in the dataset across the two years. Unfortunately, student ids are unique by year, and thus students cannot be identified across year. This concern is eliminated in models where we focus only on the reduced sample that was randomly assigned to teachers in the second year.

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 one percent of white teachers (0.63%) teach a class that consists of 100% white students.
5. 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.
6. 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. Further, 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 non-randomly assigned sample, we did not observe any indication that the results differed significantly.

7. 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.
8. A noteworthy example of the potential problem with these restricted models can be seen in the case of black females. A result that particularly jumps off the page is the effect for black females on the clarify scale in column 12 (-0.30 SD). In other words, compared to black female students, OROS students (white males in this case) in the same classroom rate black female teachers -0.30 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 to 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. For example, on the "clarify" scale, black male students rate black female teachers lower than black female students (-0.15 SD; column 10) by nearly the same amount as black female teachers rate black male teachers higher than black male students (+0.14 SD, column 7). Generally speaking, there are very few instances in this set of results that elicit concern, as there are

relatively few positive coefficients in the table and none that are statistically significant. Still, it is important to interpret these results with caution.