Teachers' Perceptions of Students' Disruptive Behavior: The Effect of Racial Congruence and Consequences for School Suspension*

Adam C. Wright
Department of Economics
Western Washington University
adam.wright@wwu.edu

December 2016

Abstract

African-American students are considerably more likely than their white peers to be rated as disruptive by their teacher and experience school discipline, but are also much less likely to have a teacher of the same race. This paper explores whether the racial or ethnic congruence of teachers and students affects teachers' perceptions of students' disruptive behavior and has larger consequences for student suspension rates. To identify the effect of racial interactions on teacher assessments, I estimate models that include both classroom and student fixed effects. I find that African-American students are rated as less disruptive when they have an African-American teacher, whereas perceptions of white and Hispanic students' disruptiveness are unaffected by having a teacher of the same race or ethnicity. I also find that African-American students with more African-American teachers are suspended less often, suggesting the underrepresentation of African-American teachers has important implications for black-white gaps in school discipline.

JEL Codes: I21, I24, J15

Keywords: Student and teacher race and ethnicity matching, disruptive behavior, school suspension

*I would like to thank Richard Startz, Kelly Bedard, Peter Kuhn, Michael Gottfried, Jenna Stearns, the UCSB labor lunch seminar participants, members of the UCSB human capital working group, and attendees of the 2015 and 2016 AEFP annual conferences for their helpful comments and suggestions. All remaining errors are my own.

1.1 Introduction

Students of color in general and African-American students in particular disproportionately experience school discipline in the United States, which likely contributes to lagging educational achievement as school discipline typically results in a loss of instructional time.[1] A potential contributing factor to black-white differences in school disciplinary outcomes may be the underrepresentation of black teachers in schools, as a growing body of research suggests that teachers assess same-race students' behavior more favorably (Dee, 2005; Downey & Pribesh, 2004; Ehrenberg, Goldhaber, & Brewer, 1995; McGrady & Reynolds, 2013). While existing research on student and teacher racial interactions has primarily focused on the implications for the black-white achievement gap, the potential for these interactions to affect the "discipline gap" has been relatively understudied. In this paper, I use a large, nationally representative dataset to determine whether the racial or ethnic congruence of teachers and students affects teachers' assessments of students' disruptive behavior and has consequences for student suspension rates.

The data used in this study come from the Early Childhood Longitudinal Study – Kindergarten Class of 1998-1999 (ECLS-K). ECLS-K includes detailed teacher assessments of student behavioral and social-emotional skills in each wave of data collection from kindergarten to fifth grade and a measure of suspension in the eighth-grade wave. While the data contain several categories of noncognitive skills, I am primarily interested in the noncognitive skills that are most strongly associated with school suspension. I show that externalizing problem behaviors, which are comprised of disruptive and acting-out behaviors, are robust predictors of

[1] Arcia (2006); Gregory, Skiba, and Noguera (2010); McCarthy and Hoge (1987); Nichols (2004); Raffaele Mendez and Knoff (2003); Skiba, Michael, Nardo, and Peterson (2002); Townsend (2000); Wu, Pink, Crain, and Moles (1982).

school suspension and thus focus my analysis on explaining how teacher-student racial dynamics influence teachers' assessments of these behaviors. Although teachers are not randomly assigned to students, the panel nature of the data along with teacher and student identifiers allow me to estimate the effect of same-race teachers on teacher assessments using both within-student and within-classroom variation. This identification strategy allows me to control for student- and classroom-specific factors that might otherwise bias my results. Estimates of the same-race effect may still be biased if, for example, students who are motivated to improve their behavior sort into classrooms with same-race teachers. I test for this threat to identification using a set of student observable characteristics that are plausibly correlated with unobserved student motivation or ability and find no evidence of problematic sorting.

Using my within-student identification strategy, I find that teachers' evaluations of African-American students' externalizing problem behaviors improve significantly when they move from a different-race teacher to a same-race teacher. I combine within-student identification and within-classroom identification, which additionally compares race-matched students' assessments to the average assessment in their classroom, and find that assessments of African-American students' externalizing behavior improve by about 0.24 standard deviations when rated by African-American teachers, an improvement equal to roughly 50% of the overall black-white gap. I find no corresponding effect of having a same-race teacher for Hispanic or white students. Robustness checks reveal that the results are entirely driven by boys and are not explained by improvements in math or reading scores. I design additional tests to assess whether the results are consistent with improvements in student behavior or merely improvements in teacher perceptions of behavior, though both of these cases might lead to less school discipline for the student. I find no evidence that previously race-matched African-American students are

rated better by subsequent different-race teachers, and thus cannot reject the hypothesis that better ratings of behavior only reflect teacher race-based perceptions.

Do these improvements in teacher perceptions of behavior translate into fewer incidences of school discipline? Identifying the causal effect of teacher-student race matching on suspension is more difficult; suspension data is given at only one point in time and measures whether a student has been suspended anytime from kindergarten through eighth grade, therefore I cannot test whether a student's likelihood of suspension changes when he moves from a different-race to a same-race teacher. Alternatively, I relate a student's *total exposure* to same-race teachers from kindergarten to eighth grade to the probability of suspension, comparing students who enter the same school in kindergarten and controlling for a rich set of student and teacher characteristics. Using this design, I show that greater exposure to same-race teachers leads to a decrease in the likelihood of suspension for African-American students. Specifically, a 30 percentage point (one standard deviation) increase in exposure to African-American teachers is associated with a 10.5-14.0 percentage point (28-38%) reduction in the probability of being suspended by eighth grade for African-American students. This effect size suggests that doubling the exposure of African-American students to African-American teachers (from 30% to 60% of the time) would shrink the black-white suspension gap by 44-59%. This study contributes to the growing literature that finds teachers tend to rate the behavior of students of their own race more favorably, but it is the first of these studies to demonstrate teacher-student race matching also has significant implications for school discipline.

This topic is of particular importance given that African-American students experience considerably higher rates of school discipline than either white or Hispanic students: 16% of African-American students experienced an out-of-school suspension during the 2011-12 school

year, compared to 5% of white students and 7% of Hispanic students (Losen et al., 2015). Even after controlling for socioeconomic indicators, students of color are overrepresented among those suspended (Skiba et al., 2005). Prior research posits that cultural mismatch, implicit bias, or negative expectations in classrooms and schools may contribute to the racial discipline gap since many teachers and schools tend to espouse white, middle-class standards of classroom deportment and behavior (Boykin, Tyler, & Miller, 2005; Morris, 2005).[2] There is some evidence that subjective interpretations may play a role in the racial gap in disciplinary outcomes, as white students are more likely to be referred to the office for observable, objective offenses (e.g., vandalism, smoking, or leaving without permission), whereas black students are more likely to be referred for behaviors requiring subjective evaluations (e.g., defiance, excessive noise, or disrespectfulness) (Gregory & Weinstein, 2008; Skiba et al., 2002).

Growing interest in how student and teacher racial interactions affect teachers' subjective evaluations of students' behavior and school discipline has led to a number of recent studies. Kinsler (2011) uses one year of North Carolina data on sixth and ninth graders to show that African-American students with white teachers are no more likely to receive an office referral than African-American students with African-American teachers within the same school. Whether these results scale to a national level or extend to another group who is underrepresented in the teacher work force but overrepresented in student suspension data – Hispanics – are contributions of this study. Bradshaw, Mitchell, O'Brennan, and Leaf (2010) examine younger cohorts, one year of data from 21 elementary schools, and find that relative to white students, African-American students are not significantly more likely to receive an office disciplinary referral in classrooms with white teachers than classrooms with African-American

[2] See Gregory et al. (2010) for a review of this literature.

teachers. The authors control for the teachers' assessments of students' disruptive behavior in their analysis (which they show is highly correlated with office referrals), so their null finding may reflect that any effect of same-race teachers on office referrals is explained by changes in perceptions of disruptive behavior.[3]

Evidence that the racial match between teachers and students affects teachers' assessments of disruptive behavior has been found in several contexts. Using data from the National Education Longitudinal Study of 1988 (NELS:88), Dee (2005) finds that eighth grade students who did not share the same race of their teacher were more likely to be labeled as disruptive and inattentive. Similarly, examining tenth grade data from the Educational Longitudinal Study of 2002, McGrady and Reynolds (2013) find that white teachers rate African-American and Hispanic students as less attentive than white students. Analyzing kindergarten data from ECLS-K, Downey and Pribesh (2004) show African-American students are rated by their teachers as exhibiting more externalizing behavior than white students on average, but when teacher race is taken into account, African-American students with African-American teachers are rated as having fewer behavioral problems than white students rated by white teachers.

Teacher ratings of student academic performance and future educational attainment also appear to be influenced by racial dynamics in the classroom. Ouazad (2014) uses ECLS-K to show that conditional on objective assessments, teachers assess same-race students in kindergarten through 5th grade more favorably in math and reading. Using tenth grade data from NELS:88, Ehrenberg, Goldhaber, and Brewer (1995) examine a composite scale that includes items about students' ability to work hard and chances of going on to college. They find that

[3] The authors do not report whether racial interactions affect the teachers' assessments of disruptive behavior.

relative to white teachers, Hispanic and African-American teachers rate students of their same race or ethnicity more positively. In a related study, Gershenson, Holt, & Papageorge (2015) show that non-black teachers have significantly lower educational attainment expectations of black students than black teachers.

Also related to this paper are studies that examine the effect of student and teacher race matching on academic achievement. Relying on data from Tennessee's Project STAR, Dee (2004) finds that African-American and white students randomly assigned to teachers of their own race have higher mathematics and reading test scores than students taught by teachers whose race differs from their own.[4] Other evaluations of teacher and student racial interactions generally confirm these positive same-race effects on student academic outcomes (Clotfelter, Ladd, & Vigdor, 2007; Egalite, Kisida, & Winters, 2015; Fairlie, Hoffman, & Oreopoulos, 2014).[5]

Two aspects of ECLS-K allow me to contribute to this literature. First, the longitudinal structure of the data allow me to use a within-student and within-classroom identification strategy to determine the effects of racial congruence on teachers' perceptions of students' disruptive behavior. Prior studies have used within-student variation to identify same-race effects on subjective teacher assessments, but they generally do not also control for unobserved classroom or teacher characteristics such as certain teachers systematically giving students better assessment scores.[6] Failure to control for these differences across classrooms would lead to biased estimates of the same-race effect if a teacher's average assessment is correlated with

[4] Chetty et al. (2011) use the STAR data to analyze the long-term impacts of early childhood education and find a positive but statistically insignificant effect of having a same-race teacher on earnings.
[5] An exception to this Howsen and Trawick (2007), who use cross-sectional data on Kentucky students in third grade and find no effect of teacher-student race match on student achievement.
[6] Figlio and Lucas (2004) find that some teachers give higher average grades regardless of student characteristics. Ouazad (2014) employs models with both student and teacher fixed effects but analyzes teacher perceptions of student math and reading ability rather than behavior.

assignment to a same-race student.[7] Second, the data contain information on school suspension, which I show is strongly correlated with externalizing behavior. This allows me to test whether teacher-student race match, beyond just affecting teachers' perceptions of behavior, impacts the likelihood of students experiencing school discipline.

The remainder of this paper is organized as follows. Section 1.2 describes the data and explores the relationship between disruptive behavior and school discipline. Section 1.3 outlines the empirical strategy and describes tests for student sorting. Section 1.4 reports results, robustness checks, and tests for possible mechanisms. Section 1.5 concludes.

1.2 Data

1.2.1 Sample Description

The data for the analysis come from the Early Childhood Longitudinal Study – Kindergarten Class of 1998-1999 (ECLS-K). Created by the National Center for Education Statistics (NCES), ECLS-K follows a nationally-representative sample of more than 20,000 kindergarten students from fall of kindergarten through eighth grade, collecting data through student assessments as well as parent, teacher, and school administrator surveys. Roughly 1,000 schools participated. Students were surveyed in six waves: fall kindergarten, spring kindergarten, spring first grade, spring third grade, spring fifth grade, and spring eighth grade. ECLS-K used a three-stage stratified sampling strategy in which geographic region represented the first sampling unit, public and private schools represented the second sampling unit, and students stratified by race and ethnicity represented the third sampling unit. Hence, the sample of children in ECLS-K

[7] Ouazad (2014) finds that being assessed by a same-race teacher is negatively correlated with the teacher's average math and reading assessments.

reflects many different types of schools and socioeconomic levels as well as different racial and ethnic backgrounds. For this study, I use the restricted version of the data.[8]

The first set of outcomes I analyze are five teacher-reported assessments of noncognitive skills measured in the spring of kindergarten through the spring of fifth grade: externalizing problem behaviors, internalizing problem behaviors, interpersonal skills, approaches to learning, and self-control.[9] These measures are adapted from the widely used Social Skills and Rating System (Gresham & Elliot, 1990), and have high test-retest reliability, internal consistency, and inter-rater reliability (Neidell & Waldfogel, 2010). Each skill is the average of a number of items and each item is rated on a 4-point Likert scale, from never (1) to very often (4). Thus, higher scale scores denote more frequently exhibited behaviors. The 5-item externalizing problem behaviors scale assesses the frequency a child argues, fights, gets angry, acts impulsively, and disturbs ongoing activities. The majority of the analysis focuses on this outcome as I demonstrate in Section 1.2.2 that externalizing behavior, more than any other student outcome, strongly correlates with school suspension.[10] The 4-item internalizing problem behaviors scale measures the extent that the child exhibits anxiety, loneliness, low self-esteem, and sadness. The 5-item interpersonal skills scale measures the frequency a child gets along with others, forms and maintains friendships, helps other children, shows sensitivity to the feelings of others, and expresses feelings, ideas, and opinions in positive ways. The 6-item approaches to learning scale rates the frequency that the child keeps his or her belongings organized, shows eagerness to learn new things, adapts to change, persists in completing tasks, and pays attention. Lastly, the 4-item

[8] See http://nces.ed.gov/ecls/ for more information.
[9] Teacher assessments of noncognitive skills are not collected in eighth grade.
[10] Additionally, improvements in externalizing behavior have been shown to benefit both labor market and health outcomes and the combined evidence from the economics and psychology literature suggest that improving these behaviors during childhood reduces crime. For a review of this literature, see Heckman, Pinto, and Savelyev (2013).

self-control scale measures the extent that the child is able to control his or her temper, respect others' property, accept his or her peers' ideas, and handle peer pressure.

I complement the teacher assessments of behaviors and skills with a measure of school discipline collected in eighth grade: a parent-reported indicator for the child ever having received an in- or out-of-school suspension.[11] Suspensions typically result in missed instructional time and have been linked with academic underperformance (Arcia, 2006; Davis & Jordan, 1994), delinquency (Balfanz, Byrnes, & Fox, 2015; Marchbanks et al., 2015) and lower educational attainment (Bertrand & Pan, 2011; Raffaele Mendez, 2003). Bertrand and Pan (2011) use the National Longitudinal Survey of Youth 1997 to show that, controlling for ASVAB math and reading scores, 7^{th} – 11^{th} graders that report ever being suspended were 21 percentage points less likely to graduate high school, 19 percentage points less likely to attend college, and 15 percentage points less likely to graduate college than students who were never suspended.[12]

I limit my sample to observations with nonmissing data on key background variables – student and instructor race, ethnicity, and gender – and require students to have at least one noncognitive outcome present. Students without teacher identifiers or that have teachers that lack information on basic teacher characteristics (experience and education level) are also dropped from the analysis. These restrictions result in 38,830 student-wave level observations for the analytical sample.[13] As Ouazad (2014) notes, the survey is designed such that data observations are mostly missing at random with regards to the sampling strategy. Due to significant attrition, I

[11] Specific definitions of in- and out-of-school suspensions are likely to vary by school. The U.S. Department of Education Office of Civil Rights defines in-school suspensions as when "a child is temporarily removed from his or her regular classroom(s) for at least half a day but remains under the direct supervision of school personnel" and out-of-school suspensions as "an instance in which a child is temporarily removed from his/her regular school for disciplinary purposes to another setting" (U.S. Department of Education Office for Civil Rights, 2014b).
[12] ASVAB stands for Armed Services Vocational Aptitude Battery. It is an aptitude test used to determine qualification by the United States Military.
[13] To comply with NCES reporting standards, sample sizes are rounded to the nearest ten.

use panel weights provided by ECLS-K to estimate representative effects. I address the issue of potential nonrandom sample attrition and how this may lead to underreporting suspensions in Section 1.4.4.

Descriptive statistics for the analytical sample are given in Table 1.1. Panel A reports student and teacher shares by race and ethnicity. Student's race and ethnicity is designated by NCES based on parent and school reports and teachers' race and ethnicity is self-reported. Students and teachers are placed in one of five mutually exclusive race and ethnicity categories: "Hispanic, any race," or the non-Hispanic categories of white, African-American, Asian, or "other race." The last category consists of American Indians, Pacific Islanders, and any non-Hispanics reporting more than one race.[14] Students are designated as having a same-race teacher if they are both Hispanic (any race) or share the same race (non-Hispanic). Panel B reports the percent of teacher-student race match by student race and ethnicity. White students have a same-race teacher 95% of the time in the sample, compared to 32% for African-American students and 25% for Hispanic students. Due to small same-race teacher sample sizes for other student race groups, my analysis focuses on these three groups.[15] Panel C gives mean student outcomes by race and ethnicity. All assessment outcomes are scaled by grade (i.e., assessment wave) to be mean zero and have a standard deviation of one in the weighted sample after the sample restrictions are applied.[16] African-American students have worse average scores for every outcome compared to white and Hispanic students. Notably, 37% of parents of African-

[14] Results are robust to alternative designations for the multiracial students, such as including them in each race category reported.
[15] I use the full sample of students to identify classroom fixed effects but I show in Section 4.3 that my results are robust to subsampling African-American, white, and Hispanic students and teachers.
[16] This might be problematic if standard deviations of assessment scores are not stable across grades. Fortunately, standard deviations tend to not vary much (e.g., the standard deviations for externalizing problem behaviors for kindergarten, first, third, and fifth grade 0.64, 0.65, 0.65, and 0.61, respectively). Results are robust to standardizing scores across all grades.

American students report that their child has received an in- or out-school suspension by eighth grade, compared to just 13% for white and 15% of Hispanic students. Suspension data are collected from parent interviews but are similar to national administrative data that report 29% of African-American and 9% of white K-12 students received an in- or out-of-school suspension during the 2011-12 school year (U.S. Department of Education Office for Civil Rights, 2014a).[17]

1.2.2 Externalizing Behavior and School Suspension

Although the data provide a rich set of student outcomes to analyze, I am most interested in the noncognitive skills that are most strongly correlated with school suspension. I therefore regress suspension on all the aforementioned student outcomes, by grade, controlling for math and reading test scores and a number of student characteristics, including variables intended to capture parental inputs.[18] Table 1.2 presents the results of these regressions. The most robust correlate of suspension is externalizing behavior.[19] There is also evidence that self-control in third grade and interpersonal skills in fifth grade relate to suspension. Interestingly, I find virtually no relationship between math or reading test scores and suspension. These results motivate my focus on analyzing externalizing behavior.

The student outcomes by race and ethnicity in Table 1.1 reveal striking differences in average externalizing behavior assessments and suspension rates between African-Americans

[17] National administrative data on public schools from the U.S. Department of Education only report suspension in each year (i.e., not whether the student has ever been suspended). Out-of-school (but not in-school) suspension data from 2011-12 are available for a sample of K-8 public schools. These data show 16% of African-American students and 5% of white students received an out-of-school suspension (Losen et al., 2015).

[18] The parental inputs are based on home-life indices adapted from Bertrand and Pan (2013) that measure in kindergarten the extent to which parents foster learning environments (the HOME index), are emotionally supportive (the WARMTH index), and use a harsh discipline style (the HARSH index). Each index is turned into indicator variable: being above the sample median for the HOME and WARMTH indices and displaying at least one harsh discipline style (e.g., the parent spanks or yells at child) for the HARSH index.

[19] I test whether the relationship between externalizing behavior and suspension differs by teacher-student race-match status in Appendix Table 1.1 and find no evidence of a difference.

and white students: African-American students are suspended nearly three times as often and have a disruptive behavior index that is 0.44 standard deviations higher on average. However, it is unclear whether these gaps are due to racial differences or simply reflect demographic differences between races. Figure 1.1 explores the extent to which these gaps can be explained by student characteristics. Panel A plots the raw mean values of externalizing behavior and suspension by student race (African-American, white, and Hispanic) by grade, revealing gaps that begin in kindergarten and persist. Panels B through D examine the regression-adjusted black-white gap in these outcomes. The regressions in Panel B control for following student characteristics: student gender, race, age at assessment, age-squared, gender-specific birthweight, and indicators for ELL status, child being in fair/poor health, attending Head Start, region, and urbanicity. These controls explain little of the black-white gaps in suspension and externalizing behavior. Adding controls for family characteristics, namely indicators for socioeconomic status quintile (based on parents' income and education level) and having both biological parents at home, in Panel C reduces black-white gaps by over 30%, but differences in externalizing behavior and suspension rate remain large. Lastly, there is little effect of additionally including parental input variables and indicators for parents' education expectations for the child, as evidenced in Panel D. Overall, the large black-white gaps in disruptive behavior and suspension rate do not appear to be simply attributable to differences in observable student characteristics.

1.2 Estimation Strategy

To assess the effect of having a same-race teacher on student noncognitive skills, I estimate a student fixed-effects model of the form:

(1.1) $$y_{ijt} = \alpha_0 + \alpha_1 RaceMatch_{ijt} + X'_{jt}\beta + \lambda_i + \varepsilon_{ijt}$$

where y_{ijt} is the assessment of student *i* by teacher *j* in year (wave) *t*. The vector X_{jt} contains teacher characteristics (gender, race, education level, experience, experience-squared) and λ_i is a student fixed effect. Student fixed effects control for time-invariant unobserved student quality and allow each student to serve as his own counterfactual.[20] Such a design controls for potential confounding factors such as overall better students sorting into classrooms with teachers of their own race or ethnicity.[21]

The variable of interest is $RaceMatch_{ijt}$, which takes the value one if student *i* and teacher *j* share the same race or ethnicity and is zero otherwise. I also decompose this variable into race-specific matches (e.g., white student with white teacher, black student with black teacher, etc.). Including student fixed effects means that the variation used to identify the coefficient on $RaceMatch_{ijt}$ comes from within a student over time. In other words, α_1 measures the effect of a change in the outcome variable associated with a change in teacher-student race-match status and is only identified for students who experience both a same-race and different-race teacher. Important for this identification strategy, a large number of minority students experience both conditions at some point between kindergarten and fifth grade: 42% of African-American students, 33% of Hispanic students, and 13% of white students change same-race teacher designations.[22] Lastly, ε_{ijt} is a stochastic error term clustered at the class level.[23]

In my preferred model, I include classroom fixed effects, γ_c, and drop the multicollinear teacher characteristics from equation (1), which can be represented as

[20] If previously race-matched students' behavior improves and carries over to a subsequent different-race teacher, then including λ_i would attenuate my estimates of α_1. I test for this in Section 4.3 and find no supporting evidence for this theory.
[21] Clotfelter, Ladd, and Vigdor (2005) provide evidence of nonrandom sorting of students to teachers.
[22] Although white students switch designations a smaller percent of the time than black or Hispanic students, they comprise a much larger share of student observations.
[23] Clustering by student or school produces very similar standard errors.

(1.2) $$y_{ic} = \alpha_0 + \alpha_1 RaceMatch_{ic} + \gamma_c + \lambda_i + \varepsilon_{ic}.$$

Here, the indices *j* and *t* are combined to create a single classroom index *c*. Including classroom fixed effects controls for unobserved differences in teacher quality and implicitly standardizes evaluation practices across classrooms as assessments of same-race students are compared to the average assessment within a classroom. Estimating this two-way fixed effects model by ordinary least squares (OLS) is computationally infeasible with a large number of students (11,680) and classrooms (13,600), and thus I rely on recent econometric advancements in estimating high-dimensional fixed effects by Guimares and Portugal (2010) and Gaure (2010).[24]

While my preferred specification addresses many issues to identifying the effect of same-race matching, the estimate of α_1 in equation (2) may be biased if time-varying unobserved student quality is correlated with both teacher-student race match and student outcomes. For example, students that are more or less motivated or likely to change their behavior may end up with a same-race teacher, perhaps because of changes to their family life. I examine this threat to validity by testing whether race-matched students have different observable characteristics that are plausibly correlated with time-varying unobserved student ability/motivation relative to non-race-matched students of the same race and in the same school and grade. Formally, I model student characteristic x_{ircsg} as

(1.3) $$x_{ircsg} = \pi_0 + \pi_1 RaceMatch_{ic} + V_c'\psi + \omega_{rsg} + u_{ircsg}\ ,$$

where students are indexed by *i*, student race/ethnicity by *r*, classrooms by c, schools by *s*, and grades by g. The vector V_c contains a set of indicators for teacher race and ω_{rsg} is a school grade by race fixed effect. The coefficient π_1 therefore tests whether students of the same race and in

[24] Specifically, the STATA command used to estimate my preferred specification is "reghdfe."

the same school grade are significantly different (along trait x) based on whether they have a same-race teacher.

1.4 Results

1.4.1 Evidence against Problematic Sorting

I first test whether student sorting may bias my main results by estimating π_1 from equation (1.3). I report estimates of the overall (pooled) race-match indicator and race-specific match indicators in Table 1.3. I examine characteristics that are likely correlated with unobserved time-varying student ability: family is in the top two SES quintiles, male student, student age, both biological parents are at home, and high parental inputs (measured at kindergarten). I find no evidence of sorting for my pooled race-match estimates, given in the first row of Table 1.3. Similarly, there is little evidence of sorting when looking at the race-specific match indicators. The exception to this is that race-matched Hispanic students appear to be different along SES measures than their non-race-matched counterparts; Hispanic students with Hispanic teachers are about 10 percentage points less likely (significant at the 10% level) to be in a high SES category than Hispanics students with non-Hispanic teachers within same school grade. However, to the extent that this represents negative sorting of Hispanics to same-race teachers, this should only serve to attenuate any positive effects of race matching for Hispanic students.[25] Thus, sorting on unobservables is unlikely to pose a serious threat to identifying the effect of race match on teacher assessments in equation (1.2).

[25] Better Hispanic students sorting into classrooms with African-American and white teachers would also attenuate any positive race-matching effects for African-American or white students in my classroom fixed effects model.

1.4.2 Same-Race Teachers and Assessments of Noncognitive Skills

Estimates of the race-match indicator from equations (1.1) and (1.2) for externalizing behavior, internalizing behavior, and approaches to learning are given in Table 1.4. When analyzing externalizing behavior, for example, the coefficient on "Race match" in models with student fixed effects would be less than zero if students are rated as being better behaved when they have a teacher of their own race compared to when they have a teacher of a different race. Results from the preferred specification with classroom and student fixed effects are listed in column (3). Additionally, column (1) reports results from a model that only includes student and teacher controls and column (2) estimates equation (1) with student fixed effects and teacher controls.

There is a significant effect of teacher-student race match on teacher assessments of externalizing behavior for African-American students. This effect is robust to specification choice and suggests that assessments of African-American students' externalizing behavior improve by about 0.24 standard deviations (in the preferred model) when they have an African-American teacher, over 50% (0.24/0.44) of the average black-white gap in externalizing behavior. There appears to be no corresponding effect of having a same-race teacher for white or Hispanic students. There is some evidence that internalizing problem behaviors and interpersonal skills improve for race-matched African-American students, but the estimates appear to be sensitive to specification choice. White students, on the other hand, appear to be judged as exhibiting more internalizing behavior when race-matched.

Estimates of the race-match indicator for student approaches to learning and self-control are listed in Table 1.5. For no race or ethnicity do I detect evidence of improvements in approaches to learning. The teacher questionnaire regarding student's self-control contains many

similar items to that of the externalizing behavior questionnaire, therefore it is not surprising that African-American teachers also tend to assess African-American students' self-control more favorably. The similarity of the externalizing behavior and self-control results (both in their magnitude and in their relation to the black-white gap in the respective scores) is a positive indication of the within-teacher consistency of the assessments. Compared to the other teacher assessments, externalizing problem behaviors are the most robust predictor of school suspension and most strongly affected by assignment to a same-race teacher. I therefore concentrate the rest of my analysis on teacher assessments of externalizing problem behaviors.

1.4.3 Same-Race Teachers and Externalizing Behavior: Robustness Checks and Mechanisms

This section provides a number of robustness checks and explores possible mechanisms driving the above results. An important consideration for understanding the relative improvements in perceptions of externalizing behavior for race-matched African-American students is what specific teacher-student racial interactions lead to these gains. My preferred specification with classroom fixed effects has the advantage of controlling for unobservable classroom factors, but restricts analysis to race-matched students relative to non-matched students in the same classroom. To estimate all teacher-student racial interactions, I drop the classroom fixed effects and add teacher controls – essentially estimating equation (1.1) – with the same-race category left out for reference. Table 1.6 reports these estimates. Each coefficient is the effect on externalizing behavior of having a teacher of a different race relative to having a same-race teacher. Both white and Hispanic teachers give worse assessments of African-American students' externalizing behavior than African-American teachers. African-American

teachers, on the other hand, do not give worse assessments of white or Hispanic students than teachers of their same race.[26] That assessments of Hispanic students' behavior do not appear to be affected by racial interactions may in part explain why Hispanic students' school disciplinary rates and levels of disruptive behavior are closer to those of white students than African-American students, despite the relative dearth of both Hispanic and African-American teachers. Furthermore, I find no evidence that these effects dissipate with teacher experience, as would be the case if race-based perceptions of behavior were due to unfamiliarity with the behavioral norms of different cultures.[27]

Since I only have enough power to test for the effects of racial congruence for African-Americans, whites, and Hispanics, I want to be sure that students and teachers of other races in my sample are not driving the results. I therefore run my preferred specification on the subsample of African-American, white, and Hispanic teachers and students. These estimates are given in Table 1.7, with the estimates from the full sample from column (3) of Table 1.4 provided for reference. The estimated African-American race-match effect for this subsample is about 30% larger than the effect from the full sample.

Previous analysis of ECLS-K data has revealed large differences in externalizing behavior between boys and girls (Bertrand & Pan, 2013). Indeed, boys "act out" about 0.45 standard deviations more than girls on average in my sample. The perception of boys' behavior may therefore be particularly sensitive to having same-race teacher given they simply have more room for improvement. I test this possibility in the last two columns of Table 1.7, where I

[26] Furthermore, the hypothesis that white teachers rate black students no different than black teachers rate white students can easily be rejected (p-value = 0.009). The hypothesis that black teachers rate Hispanic students no different than Hispanic teachers rate black students can also be rejected, but only at the 10% level (p-value = 0.056).
[27] Interacting "inexperienced" and "experienced" teacher indicators with the white teacher – black student and Hispanic teacher – black student indicators, I cannot reject the equality of the inexperienced and experienced interactions. These findings are consistent whether experienced is defined as having at least one, three, or five years of experience.

estimate my preferred specification by gender. These results suggest that the perceived improvements in disruptive behavior for African-American students with African-American teachers is entirely driven by improvements for boys, as there appears to be no improvement for African-American girls. The estimated effect for boys is large: 0.57 standard deviations. Relative to the overall black-white gap in boys' externalizing behavior (0.42 standard deviations), this estimate suggests that black boys with black teachers are assessed as less disruptive than the average white boy.

Further stratification of the sample by region in Table 1.8 reveals that the effect of teacher-student racial match is concentrated in the South, with race-matched African-American students experiencing a 0.36 standard deviation improvement in their teacher-assessed externalizing behavior. I find no statistically significant effect in other regions, though the relative imprecision of these estimates is likely due to the vast majority (73%) of racial matching for African-Americans occurring in the South.

Next, I explore possible alternative explanations for the estimated effects described above. Previous research has indicated that African-American students improve along cognitive measures when matched with African-American teachers (e.g., Dee, 2004). An important question is therefore whether race-matched African-American students improve academically when matched with African-American teachers in my sample and, if so, whether these improvements can explain African-American teachers' better perceptions of African-American student behavior. I re-run my preferred specification in equation (1.2) with math and reading test scores (scaled to mean 0 and standard deviation 1 within each wave) given in the ECLS-K which are conducted by external assessors and conform to national and state standards.[28] Results in

[28] Included in these regressions is the sample of students used to analyze externalizing behavior that have a valid math or reading test score. See Ouazad (2014) for a thorough description of the math and reading tests.

Appendix Table 1.2 do indicate that race-matched African-Americans marginally improve in math (estimates are significant at the 10% level), though I detect no effect on reading scores. Can these improvements explain my previous results? To test this, I control for student math and reading test scores and re-estimate equation (1.2) for externalizing behavior. The results in Table 1.9 indicate that cognitive improvements were not driving the results. Estimates in Table 1.9 are very similar to those given in Table 1.7, with the exception of the subsample of boys where the effect of race match is even stronger.

Another possibility is that the observed positive effects for race-matched African-American students represent more than just differences in teacher perceptions. If an African-American student's behavior is improving when he has an African-American teacher in some objective sense, then perhaps this improvement is also reflected in subsequent evaluations of the student by a teacher of a different race. To test this, I examine whether previously race-matched students (i.e., matched in the previous data collection wave) are assessed as being better behaved by different-race teachers.[29] I modify my preferred specification by including an indicator for being previously race matched and an interaction term for being both currently and previously race matched. The coefficient on the indicator for being previously race matched measures whether different-race teachers assess previously race-matched students more favorably. Because this model requires race-match data from the previous assessment wave, I analyze the sample of only first, third, and fifth grade students (i.e., kindergarten is excluded from the sample). The first column of Table 1.10 reports estimates from equation (2) on this new sample for comparison. Note that the effect of race matching for African-American students is considerably larger, perhaps suggesting that the effect of race match on teachers' perceptions of behavior is

[29] Though this definition of previous match is imperfect due to gaps in data collection (in grades two and four), I see similar results when just examining kindergarten and first grade.

stronger for later grades. The previous-race-match term and the interaction term are added in second column. Previously race-matched African-American students do not appear to be any better assessed by different-race teachers, suggesting any "real" improvements in behavior from being previously race-matched are not detected (or not detectable) by subsequent different-race teachers. Thus, I cannot reject that improvements in teacher assessments of externalizing behavior are due solely to differences in teacher race-based perceptions.

Lastly, I look for evidence of leniency towards disruptive behavior on the part of African-American teachers. If African-American students tend to act out more than their non-black peers and African-American teachers are more tolerant of disruptive behavior than white or Hispanic teachers, then my race-match results may just be a reflection of this. I investigate this by regressing students' externalizing problem behavior scores on teacher race while controlling for student and teacher characteristics and school fixed effects. The estimates provided in Table 1.11 suggest that there is little overall difference in how black, Hispanic, and white teachers assess students' disruptive behavior.

1.4.4 Same-Race Teachers and School Suspension

Does exposure to a same-race teacher have consequences for school discipline? I have shown that African-American students are considered less disruptive by African-American teachers, but this would only translate into school discipline insofar that actions measured by the externalizing problem behavior scale relate to or reflect punishable behavior. Recall that this scale measures a child's propensity to argue, fight, get angry, act impulsively, and disturb ongoing activities. While what behaviors warrant disciplinary action by a given teacher is idiosyncratic, the descriptive regressions in Table 1.2 suggest that externalizing behavior is

closely associated with receiving an in- or out-of-school suspension by eighth grade. Though teacher assessments of externalizing behavior are only reported for grades K-5 (compared to suspensions which span K-8), the race-match estimate from the first column of Table 10 suggests that the effects on externalizing behavior may be even larger for older students.

Because I only have one observation per student on suspension, I cannot rely on within-student variation in having a same-race teacher to identify the effects of race match on the likelihood of suspension. Instead, I measure a student's *total exposure* to same-race teachers using data from kindergarten, first, third, fifth, and eighth grade. Teacher race and ethnicity data are given for at most one teacher per student in grades K-3, whereas fifth and eighth grade contain information on up to two teachers each.[30] Therefore, I have data for up to seven teachers per student. On average, I have valid teacher race and ethnicity information for 6.4 teachers per student.

Since I am unable to have each student act as his own counterfactual (with a student fixed effect), I compare students of the same race that enter the same school in kindergarten as these students are likely to be similar along unobservable dimensions. I also control for a rich set of student and teacher characteristics measured in kindergarten to capture the influences of early childhood education experiences, family characteristics, and parental inputs. I (conservatively) choose to include controls measured in kindergarten because I do not know precisely when an observed suspension occurred between kindergarten and eighth grade. The linear probability model I estimate is given by

$$Suspension_{ircs} = \delta_0 + \delta_1 RaceMatchExposure_i + Z'_{ic}\phi + \sigma_{rs} + \nu_{ircs}\ , \tag{1.4}$$

[30] Fifth and eighth grade contain an English/reading teacher and either a math or science teacher.

where $Suspension_{ircs}$ is an indicator of ever being suspended by eighth grade for student *i* of race/ethnicity *r* in kindergarten classroom *c* of school *s*.[31] The vector Z_{ic} contains detailed student and teacher characteristics and σ_{rs} is a kindergarten school by race fixed effect. I also consider models with a kindergarten classroom fixed effect.[32] The covariate of interest, $RaceMatchExposure_i$, is the percent of same-race teachers a student has from kindergarten to eighth grade.[33]

Similar to my externalizing behavior analysis, I use ECLS-K panel weights in equation (1.4) to estimate representative effects. However, suspension data is collected in eighth grade and a large portion of students (44%) in my externalizing behavior analysis leave the sample before eighth grade or have no data on suspensions. If suspended students are more likely to leave the sample, it would lead me to underrepresent the number of suspended students and possibly affect my estimates in equation (1.4). I test this by regressing a binary variable for attrition (or missing suspension data) on race-specific kindergarten disruptive behavior, which I show in Table 1.2 is a good proxy for suspension. These estimates are given in Appendix Table 1.4 (with white students as the omitted category) and suggest that whites, Hispanics, and blacks experience similar attrition of disruptive students, with a one standard deviation increase in externalizing

[31]I also estimate a conditional (fixed effects) logit and get similar but less precise results. I prefer a linear probability model due to the ease of interpretation and the fact that estimating proper average partial effects in the conditional (fixed effects) logit model is not possible due to the distribution of fixed effects being unknown (Wooldridge, 2010, p.620).

[32] Chetty et al. (2011) find that students randomly assigned to better kindergarten classrooms experience significant improvements in long-term outcomes such as earnings and college attendance. Their results suggest that the long-run effects of kindergarten class quality are due to changes in noncognitive skills (effort, initiative, and disruptive behavior).

[33] A potential concern is that grade gaps in the data may lead to inaccurate measurements of $RaceMatchExposure_i$. The implicit assumption is that the percentage of same-race teachers a student has in the grades I observe is the same in the grades I do not observe. To see if this is reasonable, I divide my data into grades K-3 and 5th/8th and regress later-grade race match on earlier-grade race match along with student and teacher controls. These results are given in Appendix Table 1.3 and suggest that the percent of time a student is race matched is some grades is strongly predictive of being race matched in other grades. Further, the raw correlation between grades K-3 and 5th/8th race match for black students is 0.69.

behavior leading to about a 1% increase in the likelihood of attrition. Given that these estimates are small in magnitude, it is unlikely that nonrandom attrition poses a significant threat to the suspension analysis.

Results from estimating equation (1.4) are given in Table 1.12. Kindergarten classroom fixed effects are included in column (1) and kindergarten school by race fixed effects in column (2). Consistent with the externalizing behavior results, exposure to same-race teachers is only associated with changes in suspension rates for African-American students. African-American students are race-matched on average 30% of the time, and the results from columns (1) and (2) indicate that a 30 percentage point (one standard deviation) increase in exposure to African-American teachers is associated with a 10.5-14.0 percentage point reduction in the probability of being suspended by eighth grade. This represents a 28-38% decrease in the average black suspension rate of 0.37.[34] While this effect is large, it represents the effect of doubling the exposure of the average African-American student to African-American teachers.[35] In terms of the overall black-white suspension gap of 0.24, my estimates suggest that that doubling the exposure of African-American students to African-American teachers would shrink this gap by 44-59%.

I check the robustness of my suspension results in Table 1.13. I include the estimates from column (4) of Table 1.12 in the first column for comparison. The estimated effects of race matching for African-American students changes little when subsampling for African-American, white, and Hispanic teachers and students. Previous models with student fixed effects were able

[34] Measuring exposure as the number of times a student is matched with a same-race teacher yields nearly identical results.

[35] Doubling the exposure of the average black student to black teachers in my sample (assuming 6.4 teachers per student) would mean going from about two to four same-race teachers. Assuming students have 20 different teachers during grades K-8 (one each in K-5 and four each in 6-8), doubling the exposure of the average black student to black teachers would mean going from about 6.25 teachers to 12.5 same-race teachers.

to control for issues such as overall better behaved students sorting to same-race teachers. Since variation in same-race teacher exposure comes from across students in equation (1.4), student sorting of this nature may be an issue. I therefore control for each student's kindergarten externalizing behavior assessment in the last column of Table 1.13. Including this covariate attenuates the estimate of same-race exposure by about 10% but the point estimate remains sizable and significantly different from zero at the 5% level.[36] Overall, teacher race appears to have an important influence on African-American students' likelihood of suspension in addition to the effects on teachers' perceptions of disruptive behavior.

1.5 Conclusion

Using a large, nationally representative dataset, this paper presents evidence that teachers' assessments of African-American students' disruptive behavior are highly sensitive to the race of the teacher. Estimating models that contain both student and classroom fixed effects addresses many concerns of potential bias when estimating the effect of teacher-student racial interactions, and selective sorting of students to classrooms does not appear to be problematic. I find that teachers' evaluations of African-American students' disruptive behavior improve by about 0.24 standard deviations in classrooms with African-American teachers. This effect is large relative to racial differences in disruptive behavior, representing over 50% of the total black-white gap. The improvements in behavior are entirely driven by boys and are not explained by improvements in math or reading scores. Furthermore, I cannot reject the hypothesis that better behavioral assessments only reflect teachers' perceptions rather than actual

[36] I also estimate equation (3) separately by gender and find that the overall improvement in suspension rates for race-matched African-American students is entirely driven by boys (similar to the externalizing behavior results). Although statistically different from zero at conventional levels, the estimate for boys is large and imprecise. Due to small sample size issues I do not report these results.

improvements in behavior, as I find no evidence that previously race-matched black students are rated better by subsequent different-race teachers. Importantly, teachers' improved perceptions appear to have real consequences for school discipline: African-American students who are exposed to more African-American teachers are less likely to receive an in- or out-of-school suspension by eighth grade.

The conclusions in this paper should be of interest to policy makers, especially in light of pervasive disparities in school disciplinary outcomes between African-American and white students. Despite efforts by some U.S. states to improve the recruitment and retention of African-American teachers (Achinstein et al., 2010), they remain significantly underrepresented (U.S. Department of Education, 2013). My suspension results suggest that a more concerted effort to attract African-American teachers would lead to fewer incidences of school discipline for African-American students. My findings also have implications for how schools can conduct more fair reviews of student behavior when deciding whether certain actions warrant school discipline. To help ameliorate race-based misunderstandings, reviews of behavior should include an appropriate racial balance of evaluators.

This study contributes to the growing literature that finds teachers tend to rate the behavior of students of their own race more favorably, but it is the first of these studies to demonstrate teacher-student racial interactions also affect the likelihood that students face school discipline. The finding that black students are rated worse in non-black classrooms but non-black students' assessments are not affected by being with a black teacher suggests there may be net benefits to students (in terms of externalizing behavior assessments) of recruiting more black teachers. However, changing the racial composition of teachers may affect other student

outcomes, such as achievement (Dee, 2004), which deserve careful consideration before any policy recommendations aimed at improving overall outcomes can be made.

References

Achinstein, B., Ogawa, R.T., Sexton, D., & Freitas, C. (2010). Retaining teachers of color: A pressing problem and a potential strategy. *Review of Educational Research, 80*(1), 71-107. doi: 10.3102/0034654309355994

Arcia, E. (2006). Achievement and enrollment status of suspended students: Outcomes in a large, multicultural school district. *Education and Urban Society, 38,* 359–369.

Balfanz, R., Byrnes, V., & Fox, J. (2015). Sent home and put off track: The antecedents, disproportionalities, and consequences of being suspended in the 9th grade. In Losen, D.J., (Ed). *Closing the School Discipline Gap: Equitable Remedies for Excessive Exclusion*. New York: Teachers College Press.

Bertrand, M., & Pan, J. (2011). The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior. (NBER Working Paper No.17541). Cambridge, MA: National Bureau of Economic Research.

Bertrand, M., & Pan, J. (2013). The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior. *American Economic Journal: Applied Economics*, 5(1), 32-64.

Boykin, A. W., Tyler, K. M., & Miller, O. (2005). In search of cultural themes and their expressions in the dynamics of classroom life. *Urban Education, 40,* 521–549.

Bradshaw, C.P., Mitchell, M.M., O'Brennan, L.M., & Leaf, P.J. (2010). Multilevel exploration of factors contributing to the overrepresentation of black students in office disciplinary referrals. *Journal of Educational Psychology, 102*(2), 508-520. doi: 10.1037/a0018450

Chang, D.F. & Demyan, A. (2007). Teachers' stereotypes of Asian, Black and White students. *School Psychology Quarterly, 22*(2), 91-114. doi: http://dx.doi.org/10.1037/1045-3830.22.2.91

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. *The Quarterly Journal of Economics*,*126*(4), 1593-1660.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2005).Who teaches whom? Race and the distribution of novice teachers. *Economics of Education Review* 24(4): 377–392. doi:10.1016/j.econedurev.2004.06.008

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). How and Why do Teacher Credentials Matter for Student Achievement? (NBER Working Paper No. 12828). Cambridge, MA: National Bureau of Economic Research. doi: http://www.nber.org/papers/w12828.

Davis, J. E., & Jordan, W. J. (1994). The effects of school context, structure, and experiences on African American males in middle and high schools. *Journal of Negro Education, 63,* 570–587.

Dee, T.S. (2004). Teachers, Race, and Student Achievement in a Randomized Experiment. *Review of Economics and Statistics, 86(*1*),* 195–210. doi:10.1162/003465304323023750

Dee, T. S. (2005). A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?" *American Economic Review, 95(*2*),* 158–65.

Downey, D., & Pribesh, S. (2004). When race matters: Teachers' evaluations of students' classroom behavior. *Sociology of Education*, *77*(4), 267-282. doi: http://www.jstor.org/stable/3649390

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

Ehrenberg, R.G., Goldhaber, D.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), 547–61. doi: http://www.jstor.org/stable/2524781

Fairlie, R.W., Hoffman, F., & Oreopoulos, P. (2014). A community college instructor like me: Race and ethnicity interactions in the classroom. *American Economic Review*, *104*(8): 2567–2591. doi: 10.1257/aer.104.8.2567

Figlio, David N., & Lucas, M. (2004). Do high grading standards affect student performance? *Journal of Public Economics* 88(9): 1815–1834. doi:10.1016/S0047-2727(03)00039-2

Gaure, S. (2010) OLS with Multiple High Dimensional Category Dummies. Memorandum, Oslo University, Department of Economics.

Gershenson, S., Holt, S. B., & Papageorge, N. W. (2015). Who Believes in Me? The Effect of Student-Teacher Demographic Match on Teacher Expectations. (No. 9202). IZA Discussion Papers.

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

Guimaraes, P. and Portugal, P. (2010) A Simple Feasible Alternative Procedure to Estimate Models with High-Dimensional Fixed Effects. *Stata Journal, 10(4),* 628-649.

Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review, 103*(6), 2052-86.

Howsen, R. M., & Trawick, M. W. (2007). Teachers, race and student achievement revisited. *Applied Economics Letters, 14*(14), 1023-1027.

Losen, D.J. (2015). Closing the School Discipline Gap: Equitable Remedies for Excessive Exclusion. New York: Teachers College Press.

Losen, D., Hodson, C. I., Keith, I. I., Michael, A., Morrison, K., & Belway, S. (2015). Are We Closing the School Discipline Gap? Retrieved from http://escholarship.org/uc/item/2t36g571.pdf.

Marchbanks III, M.P., Blake, J.J., Booth, E.A., Carmichael, D., Seibert, A.L. & Fabelo, T. (2015). The economic effects of exclusionary discipline on grade retention and high school dropout. In Losen, D.J., (Ed). *Closing the School Discipline Gap: Equitable Remedies for Excessive Exclusion*. New York: Teachers College Press.

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

McGrady, P. B., & Reynolds, J. R. (2013). Racial mismatch in the classroom: Beyond black white differences. *Sociology of Education, 86* (1), 3-17. doi: 10.1177/0038040712444857

Morris, E.W. (2005). From 'middle class' to 'trailer trash': Teachers' perceptions of white students in a predominately minority school. *Sociology of Education, 78*(2), 99-121. doi: 10.1177/003804070507800201

Neidell, Matthew, and Jane Waldfogel. (2010). Cognitive and Noncognitive Peer Effects in Early Education. *Review of Economics and Statistics* 92 (3): 562–76.

Nichols, J. D. (2004). An exploration of discipline and suspension data. *The Journal of Negro Education, 73*, 408-423.

Ouazad, A. (2014). Assessed by a teacher like me: race, gender, and subjective evaluations. *Education Finance and Policy, 9*(3), 334–372.

Raffaele Mendez, L. M., & Knoff, H. M. (2003). Who gets suspended from school and why: A demographic analysis of schools and disciplinary infractions in a large school district. *Education and Treatment of Children, 26*(1), 30-51.

Raffaele Mendez, L. M. (2003). Predictors of suspension and negative school outcomes: A longitudinal investigation. In J. Wald & D. J. Losen (Eds.), *New directions for youth development: No. 99.Deconstructing the school-to-prison pipeline* (pp. 17–34). San Francisco:Jossey-Bass.

Skiba, R. J., Michael, R. S., Nardo, A.C., & Peterson, R. L. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *Urban Review, 34*, 317–342.

Skiba, R. J., Poloni-Staudinger, L., Simmons, A. B., Feggins-Azziz, L. R., & Chung, C. G. (2005). Unproven links: Can poverty explain ethnic disproportionality in special education? *Journal of Special Education, 39*(3), 130–144. doi: 10.1177/00224669050390030101

Townsend, B.L. (2000). The disproportionate discipline of African American learners: Reducing school suspensions and expulsions. *Exceptional Children, 66(3),* 381–391

U.S. Department of Education Office for Civil Rights. (2014a). Civil rights data collection: wide ranging education access and equity data collected from our nation's public schools. Retrieved from http://ocrdata. ed.gov/.

U.S. Department of Education Office for Civil Rights. (2014b). Civil rights data collection snapshot: Early childhood collection. Retrieved from http://ocrdata.ed.gov/Downloads/CRDC-Early-Childhood-Education- Snapshot.pdf.

U.S. Department of Education, National Center for Education Statistics. (2013). Characteristics of Public and Private Elementary and Secondary Schools in the United States: Results From the 2011–12 Schools and Staffing Survey.

Wooldridge, J.M. (2010). *Econometric Analysis of Cross Section and Panel Data,* 2nd Ed. Cambridge, MA: MIT Press.

Wu, S.C., Pink, W. T., Crain, R. L., & Moles, O. (1982). Student suspension: A critical reappraisal. The Urban Review, 14, 245–303.

Table 1.1 – Descriptive Statistics

	Students		Teachers		
	Mean	SD	Mean	SD	Obs.
Panel A. Student and teacher shares by race					38,830
White, non-Hispanic	0.60	0.49	0.84	0.37	
African-American, non-Hispanic	0.11	0.31	0.06	0.24	
Hispanic, any race	0.17	0.38	0.06	0.24	
Asian, non-Hispanic	0.07	0.25	0.02	0.15	
Other race, non-Hispanic	0.05	0.22	0.02	0.14	

	Mean	SD	Obs.
Panel B. Same-race teacher by student race			38,830
Overall	0.68	0.47	
White, non-Hispanic	0.95	0.22	
African-American, non-Hispanic	0.32	0.46	
Hispanic, any race	0.25	0.43	
Asian, non-Hispanic	0.10	0.30	
Other race, non-Hispanic	0.42	0.49	

	White	African-American	Hispanic	Asian	Other race
Panel C. Mean student outcomes by race					
Externalizing problem behaviors (grades K, 1, 3, 5)	-0.07	0.37	-0.07	-0.38	0.12
Observations: 38,640	(0.97)	(1.10)	(0.96)	(0.78)	(1.01)
Internalizing problem behaviors (grades K, 1, 3, 5)	-0.01	0.07	0.00	-0.21	0.10
Observations: 38,390	(0.99)	(1.07)	(1.00)	(0.86)	(0.98)
Interpersonal skills (grades K, 1, 3, 5)	0.07	-0.27	0.01	0.23	-0.15
Observations: 38,310	(0.99)	(1.04)	(0.97)	(0.91)	(0.96)
Approaches to learning (grades K, 1, 3, 5)	0.08	-0.32	-0.03	0.41	-0.09
Observations: 38,810	(0.98)	(1.03)	(1.00)	(0.86)	(0.97)
Self-control (grades K, 1, 3, 5)	0.08	-0.35	0.01	0.33	-0.16
Observations: 38,490	(0.97)	(1.06)	(0.97)	(0.86)	(0.99)
Ever suspended, measured in grade 8	0.13	0.37	0.15	0.04	0.14
Observations: 5,570	(0.34)	(0.48)	(0.35)	(0.19)	(0.34)

Notes: The "other race, non-Hispanic" category consists of American Indians, Pacific Islanders, and those reporting more than one race. All scores are standardized to be mean zero and standard deviation one within each grade. A lower value signifies a more favorable outcome for externalizing and internalizing problem behaviors. A higher value signifies a more favorable outcome for interpersonal skills, approaches to learning, and self-control. Panel A and Panel B show percentages for the unweighted data. Observations used to calculate student group means and standard deviations in Panel C are weighted using ECLS-K panel weights. Reported observations are rounded to nearest 10 to comply with NCES stipulations.

Table 1.2 – Relationship between Suspension and Assessments of Cognitive and Noncognitive Skills

	Outcome: Ever suspended, measured in Grade 8			
	Spring K	Grade 1	Grade 3	Grade 5
Externalizing problem behaviors	0.052***	0.055***	0.054***	0.083***
	(0.014)	(0.013)	(0.014)	(0.013)
Internalizing problem behaviors	0.005	-0.005	-0.004	-0.001
	(0.009)	(0.009)	(0.010)	(0.010)
Interpersonal skills	-0.0011	0.0012	0.004	-0.032**
	(0.014)	(0.015)	(0.015)	(0.013)
Approaches to learning	0.013	-0.010	-0.020	-0.001
	(0.012)	(0.013)	(0.013)	(0.013)
Self-control	-0.014	-0.023	-0.041**	-0.014
	(0.017)	(0.016)	(0.016)	(0.019)
Math test score	-0.014	0.018*	0.003	-0.001
	(0.012)	(0.011)	(0.013)	(0.013)
Reading test score	-0.004	-0.018	-0.003	0.001
	(0.011)	(0.012)	(0.014)	(0.013)
Controls				
Student	Yes	Yes	Yes	Yes
Observations	5,600	5,140	4,600	4,900
R^2	0.17	0.19	0.20	0.23

Notes: The basic sample restrictions are described in the text. The sample is further restricted to students with nonmissing suspension data, math and reading test scores, and student control variables listed below. Each column represents a separate OLS regression. A lower value signifies a more favorable outcome for externalizing and internalizing problem behaviors. A higher value signifies a more favorable outcome for interpersonal skills, approaches to learning, and self-control. Student controls include student gender, race, age at assessment, age-squared, gender-specific birthweight, indicators for the HOME, WARMTH, and HARSH indices discussed in the text, and indicators for parents' education expectations for the child, SES quintile, both biological parents at home, ELL status, child being in fair/poor health, attending Head Start, region, and urbanicity. Robust standard errors given in parentheses. Observations are weighted using ECLS-K panel weights rounded to nearest 10 to comply with NCES stipulations.

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

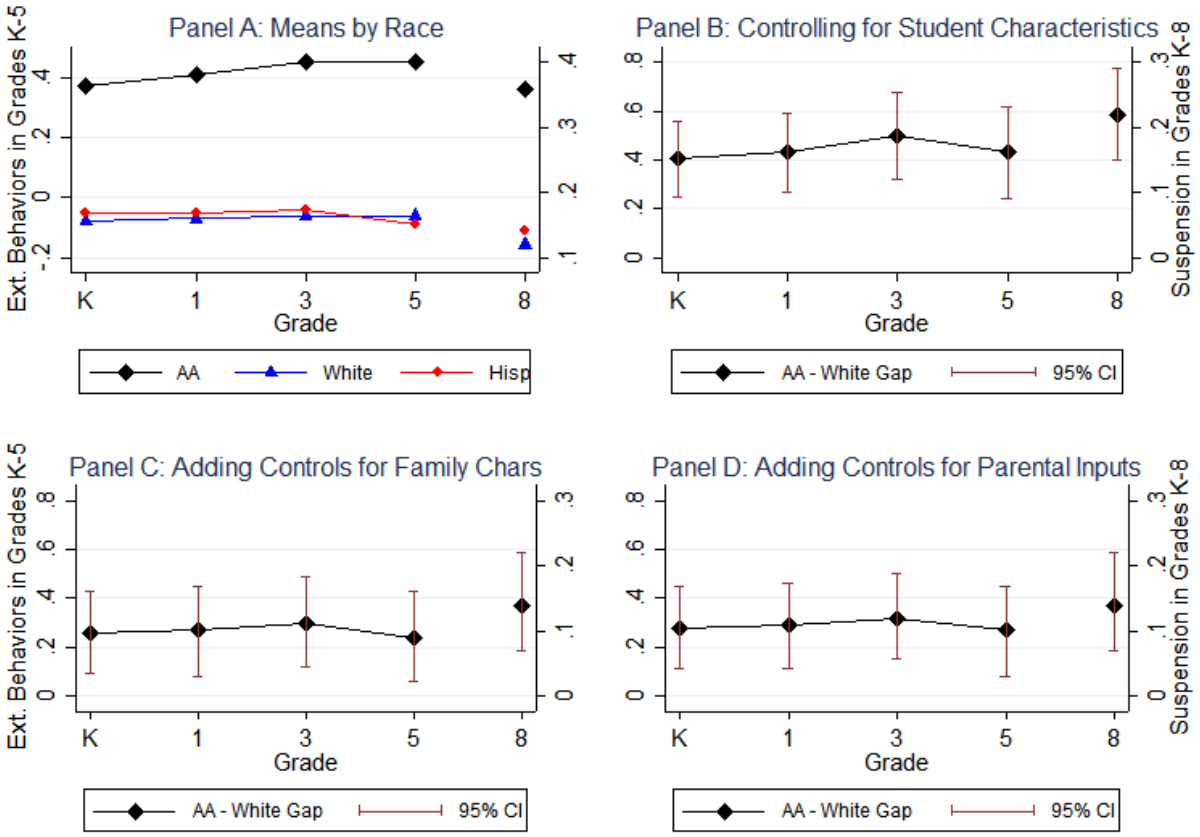

Figure 1.1: Externalizing Behavior and School Suspension by Race

Notes: The basic sample restrictions are described in the text. The sample is further restricted to students with nonmissing suspension data and student control variables listed below. A lower value signifies a more favorable outcome for externalizing behavior. Panel A plots the raw mean values of externalizing behavior and suspension by student race (African-Americans, whites, and Hispanics). Panels B through D examine the regression-adjusted gap in these outcomes between African-American and white students (with robust standard errors). Panel B controls for student gender, race, age at assessment, age-squared, gender-specific birthweight, and indicators for ELL status, child being in fair/poor health, attending Head Start, region, and urbanicity. Panel C adds family characteristics variables: indicators for SES quintile and both biological parents at home. Panel D adds the family characteristics variables plus indicators for the HOME, WARMTH, and HARSH indices discussed in the text and indicators for parents' education expectations for the child. Observations are weighted using ECLS-K panel weights.

Table 1.3 – Tests for Sorting

	Outcome					
	Student family high SES	Male student	Student age (months)	Both biological parents	High HOME index	High WARMTH index
Overall effect						
Race match	-0.009	0.030	-0.189	-0.003	0.023	-0.013
	(0.022)	(0.026)	(0.275)	(0.024)	(0.037)	(0.030)
Effect by race						
Race match: African-American	-0.039	0.008	-1.328	-0.049	0.020	-0.036
	(0.057)	(0.077)	(0.823)	(0.073)	(0.107)	(0.098)
Race match: White	0.058	-0.004	0.202	0.038	0.100	0.043
	(0.049)	(0.055)	(0.555)	(0.052)	(0.086)	(0.076)
Race match: Hispanic	-0.104*	0.105	-0.097	-0.027	-0.087	-0.100
	(0.056)	(0.071)	(0.693)	(0.074)	(0.122)	(0.098)
Fixed effects						
School-grade-race	Yes	Yes	Yes	Yes	Yes	Yes
Controls						
Teacher and student race	Yes	Yes	Yes	Yes	Yes	Yes
Observations	34,320	38,830	36,370	34,220	29,680	32,600

Notes: Each sub-heading ("overall effect" and "effect by race") represents a separate OLS regression for each outcome. Though the same-race effect for all student race categories included in each regression, I report only the three largest categories here. Standard errors clustered at the school-grade-race level and are given in parentheses. Observations are weighted using ECLS-K panel weights and rounded to nearest 10 to comply with NCES stipulations.
*p < 0.1, **p < 0.05, ***p<0.01

Table 1.4 – Estimated Effects of Student and Teacher Race Matching on Student Externalizing Behavior, Internalizing Behavior, and Approaches to Learning

	(1)	(2)	(3)
Outcome: Externalizing problem behaviors	Obs: 38,640		
Overall effect			
Race match	-0.019	-0.048	-0.041
	(0.037)	(0.035)	(0.049)
Effect by race			
Race match: African-American	-0.192*	-0.214**	-0.235**
	(0.101)	(0.102)	(0.120)
Race match: White	0.022	-0.001	-0.041
	(0.082)	(0.072)	(0.085)
Race match: Hispanic	-0.012	0.037	0.144
	(0.099)	(0.093)	(0.136)
Outcome: Internalizing problem behaviors	Obs: 38,390		
Overall effect			
Race match	-0.043	0.010	0.069
	(0.036)	(0.048)	(0.053)
Effect by race			
Race match: African-American	-0.272**	-0.158	-0.077
	(0.110)	(0.131)	(0.156)
Race match: White	0.050	0.163*	0.173*
	(0.089)	(0.090)	(0.091)
Race match: Hispanic	-0.015	-0.068	-0.001
	(0.013)	(0.015)	(0.014)
Outcome: Interpersonal skills	Obs: 38,310		
Overall effect			
Race match	0.025	0.053	0.054
	(0.040)	(0.042)	(0.048)
Effect by race			
Race match: African-American	0.165	0.102	0.246*
	(0.110)	(0.126)	(0.142)
Race match: White	-0.016	0.018	-0.065
	(0.091)	(0.087)	(0.092)
Race match: Hispanic	0.055	0.143	0.069
	(0.112)	(0.114)	(0.110)
Fixed effects			
Student	No	Yes	Yes
Classroom	No	No	Yes
Controls			
Teacher	Yes	Yes	No
Student	Yes	No	No

Notes: Each sub-heading ("overall effect" and "effect by race") represents a separate OLS regression. A lower value signifies a more favorable outcome for externalizing and internalizing problem behaviors. A higher value signifies a more favorable outcome for interpersonal skills. Though the same-race effect for all student race categories is included in each regression, I report only the three largest categories here. Teacher controls include education level, experience, experience-squared, gender, race, and ethnicity. Student controls include race, ethnicity, and gender. Standard errors clustered at the class level are given in parentheses. Observations are weighted using ECLS-K panel weights and rounded to nearest 10 to comply with NCES stipulations.
*p < 0.1, **p < 0.05, ***p<0.01

Table 1.5 – Estimated Effects of Student and Teacher Race Matching on Student Approaches to Learning and Self-Control

	(1)	(2)	(3)
Outcome: Approaches to learning	Obs: 38,810		
Overall effect			
Race match	0.026	0.045	0.026
	(0.039)	(0.039)	(0.043)
Effect by race			
Race match: African-American	0.194	0.106	0.040
	(0.104)	(0.118)	(0.124)
Race match: White	-0.069	-0.044	0.039
	(0.090)	(0.076)	(0.097)
Race match: Hispanic	0.099	0.161	0.001
	(0.103)	(0.104)	(0.107)
Outcome: Self-control	Obs: 38,490		
Overall effect			
Race match	0.023	0.075*	0.032
	(0.041)	(0.042)	(0.046)
Effect by race			
Race match: African-American	0.244**	0.206*	0.193
	(0.108)	(0.123)	(0.125)
Race match: White	-0.019	0.060	-0.018
	(0.090)	(0.089)	(0.198)
Race match: Hispanic	-0.022	-0.002	-0.030
	(0.113)	(0.124)	(0.124)
Fixed effects			
Student	No	Yes	Yes
Classroom	No	No	Yes
Controls			
Teacher	Yes	Yes	No
Student	Yes	No	No

Notes: Each sub-heading ("overall effect" and "effect by race") represents a separate OLS regression. A higher value signifies a more favorable outcome for approaches to learning and self-control. Though the same-race effect for all student race categories is included in each regression, I report only the three largest categories here. Teacher controls include education level, experience, experience-squared, gender, race, and ethnicity. Student controls include race, ethnicity and gender. Standard errors clustered at the class level are given in parentheses. Observations are weighted using ECLS-K panel weights and rounded to nearest 10 to comply with NCES stipulations.
*p < 0.1, **p < 0.05, ***p<0.01

Table 1.6 – Estimated Effects of All Race Interactions on Student Externalizing Behavior

	Race of the teacher		
	African-American	White	Hispanic
Outcome: Externalizing problem behaviors			
Race of the student			
African-American	Reference	0.273***	0.257*
		(0.079)	(0.140)
White	0.048	Reference	-0.035
	(0.095)		(0.098)
Hispanic	-0.067	-0.012	Reference
	(0.089)	(0.055)	
Fixed effects			
Student		Yes	
Controls			
Teacher		Yes	
Observations		38,640	

Notes: All estimates in this table come from the same OLS regression. Though all race interactions are included in the regression, I report only the interactions for the three largest categories here. Teacher controls include education level, experience, experience-squared, gender, race, and ethnicity. Standard errors clustered at the class level are given in parentheses. Observations are weighted using ECLS-K panel weights and rounded to nearest 10 to comply with NCES stipulations.

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

Table 1.7 – Estimated Effects of Student and Teacher Race Matching on Student Externalizing Behavior: Robustness and Mechanisms

	Full sample	AA, white, Hispanic teachers and students	Female students	Male students
Outcome: Externalizing problem behaviors				
Overall effect				
Race match	-0.041	-0.024	-0.071	-0.041
	(0.049)	(0.062)	(0.096)	(0.049)
Effect by race				
Race match: African-American	-0.235**	-0.310**	0.089	-0.573**
	(0.120)	(0.131)	(0.149)	(0.265)
Race match: White	-0.041	0.024	-0.059	0.107
	(0.085)	(0.120)	(0.116)	(0.186)
Race match: Hispanic	0.144	0.293	0.060	0.036
	(0.136)	(0.187)	(0.157)	(0.223)
Fixed effects				
Student	Yes	Yes	Yes	Yes
Classroom	Yes	Yes	Yes	Yes
Observations	38,640	33,270	19,280	19,360

Notes: Each sub-heading ("overall effect" and "effect by race") represents a separate OLS regression. Though the same-race effect for all student race categories is included in each regression, I report only the three largest categories here. Standard errors clustered at the class level are given in parentheses. Observations are weighted using ECLS-K panel weights and rounded to nearest 10 to comply with NCES stipulations.
*p < 0.1, **p < 0.05, ***p<0.01

Table 1.8 – Estimated Effects of Student and Teacher Race Matching on Externalizing Behavior, by Region

	Northeast	Midwest	South	West
Outcome: Externalizing problem behaviors				
Effect by race				
Race match: African-American	-0.239	0.320	-0.364**	0.047
	(0.209)	(0.317)	(0.157)	(0.280)
Fixed effects				
Student	Yes	Yes	Yes	Yes
Classroom	Yes	Yes	Yes	Yes
Observations	6,680	10,140	11,720	7,740

Notes: Each column represents a separate OLS regression. Though the same-race effect for all student race categories is included in each regression, I report only African-Americans here. Standard errors clustered at the class level are given in parentheses. Observations are weighted using ECLS-K panel weights and rounded to nearest 10 to comply with NCES stipulations.
*p < 0.1, **p < 0.05, ***p<0.01

Table 1.9 – Estimated Effects of Student and Teacher Race Matching on Externalizing Behavior, Controlling for Math and Reading Scores

	Full sample (with math and reading scores)	AA, white, Hispanic teachers and students	Female students	Male students
Outcome: Externalizing problem behaviors				
Effect by race				
Race match: African-American	-0.236*	-0.285**	0.078	-0.739***
	(0.131)	(0.140)	(0.160)	(0.026)
Math score	-0.049***	-0.056***	-0.035	-0.051*
	(0.016)	(0.018)	(0.025)	(0.026)
Reading score	-0.029**	-0.036**	-0.044**	-0.017
	(0.013)	(0.014)	(0.020)	(0.024)
Fixed effects				
Student	Yes	Yes	Yes	Yes
Classroom	Yes	Yes	Yes	Yes
Observations	35,610	30,700	17,800	17,810

Notes: Each column represents a separate OLS regression. Though the same-race effect for all student race categories is included in each regression, I report only African-Americans here. Standard errors clustered at the class level are given in parentheses. Observations are weighted using ECLS-K panel weights and rounded to nearest 10 to comply with NCES stipulations.
*p < 0.1, **p < 0.05, ***p<0.01

Table 1.10 – Testing Whether Different-Race Teachers Assess Previously Race-Matched Students More Favorably

	Grades 1, 3, and 5 only	
Outcome: Externalizing problem behaviors		
Effect by race		
Previous race match: African-American	-----	0.062
		(0.212)
Current race match: African-American	-0.605***	-0.595***
	(0.160)	(0.219)
Interaction: African-American	-----	0.138
		(0.217)
Fixed effects		
Student	Yes	Yes
Classroom	Yes	Yes
Observations	27,960	27,960

Notes: Each column represents a separate regression. Though the previous grade effect for all student race categories is included in each regression, I report only African-Americans here. Standard errors clustered at the class level and are given in parentheses. Observations are weighted using ECLS-K panel weights and rounded to nearest 10 to comply with NCES stipulations.
*p < 0.1, **p < 0.05, ***p<0.01

Table 1.11 – Relative Leniency in Assessing Student Behavior by Teacher Race/Ethnicity

	Outcome: Externalizing problem behaviors
Teacher black	0.039
	(0.054)
Teacher Hispanic	-0.056
	(0.040)
Student black	0.361***
	(0.044)
Student Hispanic	-0.128***
	(0.036)
Fixed effects	
School	Yes
Controls	
Teacher	Yes
Student	Yes
Observations	33,980

Notes: All estimates in this table come from the same OLS regression. Though all student and teacher race/ethnicity categories are included in the regression, I report only the three largest categories here. The omitted race category is "white." Teacher controls include gender, education category, experience, and experience squared. Student controls include gender, age, age squared, SES quintile, and an indicator for whether the student has both biological parents at home. Standard errors clustered at the class level and are given in parentheses. Observations are weighted using ECLS-K panel weights and rounded to nearest 10 to comply with NCES stipulations.
*p < 0.1, **p < 0.05, ***p<0.01

Table 1.12 – Estimated Effects of Student and Teacher Race Matching on Suspension

	(1)	(2)
Outcome: Ever suspended		
Overall effect		
Race match	-0.079	-0.078
	(0.085)	(0.097)
Effect by race		
Race match exposure: African-American	-0.352*	-0.468**
	(0.210)	(0.211)
Race match exposure: White	-0.043	0.012
	(0.140)	(0.150)
Race match exposure: Hispanic	0.197	0.056
	(0.147)	(0.157)
Fixed effects		
Kindergarten classroom	Yes	No
Kindergarten school by race	No	Yes
Controls		
Teacher	No	Yes
Student	Yes	Yes
Observations	5,570	5,570
Black race match mean (SD) = 0.30 (0.30)		
Estimated effect of 1 SD increase in same-race teacher exposure (black)	-10.50 percentage points	-14.04 percentage points

Notes: Each sub-heading ("overall effect" and "effect by race") represents a separate OLS regression. All teacher and student controls are measured in kindergarten. Though the same-race effect for all student race categories is included in each regression, I report only the three largest categories here. Teacher controls include education level, experience, experience-squared, gender, race, and ethnicity. Student controls include race, ethnicity, gender, age at kindergarten entry, age-squared, gender-specific birthweight, indicators for the HOME, WARMTH, and HARSH indices discussed in the text, and indicators for parents' education expectations for the child, SES quintile, both biological parents at home, ELL status, child being in fair/poor health, and attending Head Start. Standard errors clustered at the school-grade-race level are given in parentheses. Observations are weighted using ECLS-K panel weights and rounded to nearest 10 to comply with NCES stipulations.
*p < 0.1, **p < 0.05, ***p<0.01

Table 1.13 – Estimated Effects of Student and Teacher Race Matching on Suspension: Robustness

	Full sample	AA, white, Hispanic teachers and students	Control for K ext. problem behaviors
Outcome: Ever suspended			
Overall effect			
Race match	-0.078	-0.105	-0.076
	(0.097)	(0.102)	(0.099)
Effect by race			
Race match exposure: African-American	-0.468**	-0.472**	-0.428**
	(0.211)	(0.210)	(0.212)
Race match exposure: White	0.012	0.024	0.033
	(0.150)	(0.153)	(0.151)
Race match exposure: Hispanic	0.056	0.052	0.042
	(0.157)	(0.158)	(0.163)
Fixed effects			
Kindergarten classroom	No	No	No
Kindergarten school by race	Yes	Yes	Yes
Controls			
Teacher	Yes	Yes	Yes
Student	Yes	Yes	Yes
Student K ext. behavior	No	No	Yes
Observations	5,570	5,050	5,570

Notes: Each sub-heading ("overall effect" and "effect by race") represents a separate OLS regression. All teacher and student controls are measured in kindergarten. Though the same-race effect for all student race categories is included in each regression, I report only the three largest categories here. Teacher controls include education level, experience, experience-squared, gender, race, and ethnicity. Student controls include race, ethnicity, gender, age at kindergarten entry, age-squared, gender-specific birthweight, indicators for the HOME, WARMTH, and HARSH indices discussed in the text, and indicators for parents' education expectations for the child, SES quintile, both biological parents at home, ELL status, child being in fair/poor health, and attending Head Start. Standard errors clustered at the school-grade-race level are given in parentheses. Observations are weighted using ECLS-K panel weights and rounded to nearest 10 to comply with NCES stipulations.
*p < 0.1, **p < 0.05, ***p<0.01

Appendix Table 1.1: Relationship between Suspension and Externalizing Behavior: Controlling for Student and Teacher Race Matching

	Outcome: Ever suspended, measured in Grade 8			
	Spring K	Grade 1	Grade 3	Grade 5
Externalizing problem behaviors	0.060**	0.071***	0.032	0.089***
	(0.025)	(0.024)	(0.024)	(0.027)
Race match*externalizing problem behaviors	-0.013	-0.025	0.033	-0.010
	(0.029)	(0.029)	(0.030)	(0.030)
Controls				
Student	Yes	Yes	Yes	Yes
Observations	5,600	5,140	4,600	4,900
R^2	0.17	0.19	0.21	0.24

Notes: A lower value signifies a more favorable outcome for externalizing problem behaviors. Student controls include student gender, race, age at assessment, age-squared, gender-specific birthweight, indicators for the HOME, WARMTH, and HARSH indices discussed in the text, and indicators for parents' education expectations for the child, SES quintile, both biological parents at home, ELL status, child being in fair/poor health, attending Head Start, region, and urbanicity. Also included is a dummy for teacher-student race match. Robust standard errors given in parentheses. Observations are weighted using ECLS-K panel weights and rounded to nearest 10 to comply with NCES stipulations.
*p < 0.1, **p < 0.05, ***p<0.01

Appendix Table 1.2 – Estimated Effects of Student and Teacher Race Matching on Math and Reading Scores

	Outcome	
	Math score	Reading score
Effect by race		
Race match: African-American	0.178*	-0.058
	(0.093)	(0.118)
Fixed effects		
Student	Yes	Yes
Classroom	Yes	Yes
Observations	36,210	37,730

Notes: Each column represents a separate regression. Though the same-race effect for all student race categories is included in each regression, I report only African-Americans here. Standard errors clustered at the class level are given in parentheses. Observations are weighted using ECLS-K panel weights and rounded to nearest 10 to comply with NCES stipulations.
*p < 0.1, **p < 0.05, ***p<0.01

Appendix Table 1.3 – Race-Match Correlation between Early and Later Grades

	Outcome: Percent of time race matched 5th and 8th grade
Overall effect	
Race match % (K-3)	0.303***
	(0.031)
Effect by race	
Race match % (K-3): African- American	0.454***
	(0.083)
Race match % (K-3): White	0.205***
	(0.059)
Race match % (K-3): Hispanic	0.287***
	(0.052)
Controls	
Teacher	Yes
Student	Yes
Observations	5,570

Notes: Each sub-heading ("overall effect" and "effect by race") represents a separate OLS regression. Robust standard errors are given in parentheses. Observations are weighted using ECLS-K panel weights and rounded to nearest 10 to comply with NCES stipulations.
*p < 0.1, **p < 0.05, ***p<0.01

Appendix Table 1.4 – Predicting Sample Attrition before Suspension Data Collected in Eighth Grade

	Outcome: Attrition before Eighth Grade
K Externalizing problem behaviors	0.011*
	(0.006)
Black*K externalizing problem behaviors	0.000
	(0.014)
Hispanic*K externalizing problem behaviors	-0.005
	(0.014)
Black	0.187***
	(0.016)
Hispanic	0.140***
	(0.013)
Observations	9,930

Notes: All estimates in this table come from the same OLS regression. Though all student race/ethnicity categories (and their interaction with externalizing behavior) are included in the regression, I report only the three largest categories here. The omitted race category is "white." Also include in the regression is a student gender indicator. Robust standard errors are given in parentheses. Observations are rounded to nearest 10 to comply with NCES stipulations.
*p < 0.1, **p < 0.05, ***p<0.01