UNIVERSITY OF CALIFORNIA

Santa Barbara

Three Essays on the Economics of Education

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Economics

by

Adam Charles Wright

Committee in charge:

Professor Richard Startz, Chair

Professor Kelly Bedard

Professor Peter Kuhn

June 2016

The dissertation of Adam Charles Wright is approved.

_____

Peter Kuhn

_____

Kelly Bedard

_____

Richard Startz, Committee Chair

June 2016

Three Essays on the Economics of Education

by

Adam Charles Wright

ACKNOWLEDGEMENTS

I am truly grateful for the help and support I have received while completing this dissertation. I would especially like to thank my advisor, Dick Startz, and my other committee members Kelly Bedard and Peter Kuhn. Dick has been everything I could ask for in a mentor. His kindness, patience, and high expectations have been instrumental in my development. I am fortunate to have been his student. Thank you to Kelly for your thoughtful comments and generosity with your time. Thank you to Peter for diligently reading my work and steering my research. I also thank my two coauthors, Michael Gottfried and Vi-Nhuan Le.

My parents have provided constant support throughout my graduate studies. They have always celebrated my work, even the smallest of accomplishments. I thank my father for instilling in me the importance of selflessness and work ethic. I thank my mother for showing me what it means to have an open mind and heart. Lastly, I would like to thank my wife, Mary. Her unwavering belief in me continues to make all the difference.

VITA OF ADAM CHARLES WRIGHT
June 2016

EDUCATION

Ph.D. Economics, University of California Santa Barbara, 2016
M.A. Economics, University of California Santa Barbara, 2011
M.Ed. Secondary Education, Arizona State University, 2010
B.S. Mathematical Economics (Honors), University of Richmond, 2008

RESEARCH FIELDS

Labor Economics, Economics of Education, Applied Econometrics, Applied Microeconomics

COMMITTEE MEMBERS

Richard Startz (Chair), Kelly Bedard, Peter Kuhn

PUBLICATIONS

"A Kindergarten Teacher Like Me: The Role of Student-Teacher Race in Social-Emotional Development," with Michael Gottfried and Vi-Nhuan Le. Forthcoming at the American Educational Research Journal.

GRANTS, AWARDS, & FELLOWSHIPS

Robert T. Deacon Graduate Fellowship in Economics, UCSB, Spring 2015
Broom Center of Demography Graduate Student Fellow, UCSB, 2012-Present
Graduate Student Association Excellence in Teaching Award nominee, UCSB, 2014
Sarah D. Barder Fellowship, Johns Hopkins University, 2014
Instructional Improvement Grant (with Richard Startz), UCSB, 2013-2014
Department of Economics Outstanding Teaching Assistant, UCSB, Fall 2012
Department of Economics Teaching Assistant of the Year, UCSB, 2011-2012

TEACHING EXPERIENCE (UCSB)

Introduction to Econometrics (Instructor), Winter 2014, Spring 2014
Intermediate Microeconomic Theory (Instructor), Summer 2013
Principles of Microeconomics (Teaching Assistant), Fall 2011, Winter 2012, Fall 2012, Winter 2013, Fall 2014, Winter 2015, Fall 2015, Winter 2016
Introduction to Econometrics (Teaching Assistant), Fall 2013
Intermediate Microeconomic Theory (Teaching Assistant), Winter 2011, Summer 2012, Spring 2013
Principles of Macroeconomics (Teaching Assistant), Spring 2012
Statistics with Business and Economics Applications (Teaching Assistant), Spring 2011

PROFESSIONAL EXPERIENCE

Human Capital Research Working Group, supervised by Richard Startz and Kelly Bedard (UCSB), 2012-2016
Constructed and implemented online qualification exam to achieve uniform minimum level of mathematical fluency for UCSB undergraduate introductory econometrics, 2013-2014
Research Assistant to Charles Kolstad (UCSB), University of California Center for Energy and Environmental Economics, 2011-2012
Institute for Training Educational Evaluation, University of California Educational Evaluation Center, July 2012
Teacher, Teach For America (Agua Fria High School, Avondale, AZ), 2008-2010
Referee for Industrial & Labor Relations Review

ABSTRACT

Three Essays on the Economics of Education

by

Adam Charles Wright

This dissertation consists of three essays in applied microeconomics that investigate how teacher and student interactions affect human capital and skill accumulation. The first essay addresses the topic of how race and ethnicity affect teachers' perceptions students' behavior. 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.

The second essay, coauthored with Michael Gottfried and Vi-Nhuan Le, examines whether gaps in social-emotional skills between students of color and white students is smaller in classrooms with teachers of color. Our nation's classrooms have become increasingly racially and ethnically diverse. Given these demographic changes, many policymakers and practitioners have expressed the need for increased attention to how teacher diversity might be linked to reducing racial/ethnic differences in teachers' ratings of social-emotional skills for students of color. Using the most recent nationally representative data, we investigated whether kindergartners have different social-emotional ratings when they had a teacher whose racial/ethnic group was the same as their own. We found that having a teacher of the same race was unrelated to teachers' ratings of children's internalizing problem behaviors, interpersonal skills, approaches to learning, and self-control. However, students whose teachers' race/ethnicity matched their own had more favorable ratings of externalizing behaviors. Results are discussed in terms of implications for school disciplinary policies.

In the third essay, I estimate spillovers of teachers from a selective alternative teacher certification program. The growing prevalence of teachers from selective alternative teacher certification programs has prompted research into how these teachers affect their own students' performance. Little is known, however, about how the presence of this type of teacher might affect the performance of other teachers. This paper explores the extent to which the presence of teachers from a selective alternative certification program, Teach For America, affects grade-level student achievement in nearby grades. Using data from California elementary schools, I find that grades adjacent to Teach For America grades improve in both math and English as Teach For America presence increases.

Chapter 1

Teachers' Perceptions of Students' Disruptive Behavior:

The Effect of Racial Congruence and Consequences for School Suspension

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

[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).

externalizing problem behaviors, which are comprised of disruptive and acting-out behaviors, are robust predictors of 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

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

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 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

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

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 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

[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.

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 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

[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.

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

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:1999). Created by the National Center for Education Statistics (NCES), ECLS-K:1999 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:1999 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:1999 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,

[8] See http://nces.ed.gov/ecls/ for more information.

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 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

[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).

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 use panel weights provided by ECLS-K:1999 to

[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.

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-

[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 have worse average scores for every outcome compared to white and Hispanic students. Notably, 37% of parents of African-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

[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.

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 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:

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

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

[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.

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

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

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

[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.

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

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

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

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 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).

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

[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.

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

[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).

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:1999 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 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

[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.

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:1999 which are conducted by external assessors and conform to national and state standards.[28] Results in 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

[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.

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 stronger for later grades. The previous-race-match term and the interaction term are added in second

[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.

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:1999 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.

[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 $5^{th}/8^{th}$ 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 $5^{th}/8^{th}$ race match for black students is 0.69.

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 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%.

[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.

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 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

[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.

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 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 (Chapter 1)

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.

Tables (Chapter 1)

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) Observations: 38,640	-0.07 (0.97)	0.37 (1.10)	-0.07 (0.96)	-0.38 (0.78)	0.12 (1.01)
Internalizing problem behaviors (grades K, 1, 3, 5) Observations: 38,390	-0.01 (0.99)	0.07 (1.07)	0.00 (1.00)	-0.21 (0.86)	0.10 (0.98)
Interpersonal skills (grades K, 1, 3, 5) Observations: 38,310	0.07 (0.99)	-0.27 (1.04)	0.01 (0.97)	0.23 (0.91)	-0.15 (0.96)
Approaches to learning (grades K, 1, 3, 5) Observations: 38,810	0.08 (0.98)	-0.32 (1.03)	-0.03 (1.00)	0.41 (0.86)	-0.09 (0.97)
Self-control (grades K, 1, 3, 5) Observations: 38,490	0.08 (0.97)	-0.35 (1.06)	0.01 (0.97)	0.33 (0.86)	-0.16 (0.99)
Ever suspended, measured in grade 8 Observations: 5,570	0.13 (0.34)	0.37 (0.48)	0.15 (0.35)	0.04 (0.19)	0.14 (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:1999 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:1999 panel weights rounded to nearest 10 to comply with NCES stipulations.
*p < 0.1, **p < 0.05, ***p<0.01

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:	-0.039	0.008	-1.328	-0.049	0.020	-0.036
African-American	(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:1999 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:1999 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:1999 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:1999 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:1999 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:1999 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:1999 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:1999 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 African-American	0.039
	(0.054)
Teacher Hispanic	-0.056
	(0.040)
Student African-American	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:1999 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:1999 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:	-0.468**	-0.472**	-0.428**
African- American	(0.211)	(0.210)	(0.212)
Race match exposure:	0.012	0.024	0.033
White	(0.150)	(0.153)	(0.151)
Race match exposure:	0.056	0.052	0.042
Hispanic	(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:1999 panel weights and rounded to nearest 10 to comply with NCES stipulations.
*p < 0.1, **p < 0.05, ***p<0.01

Figures (Chapter 1)

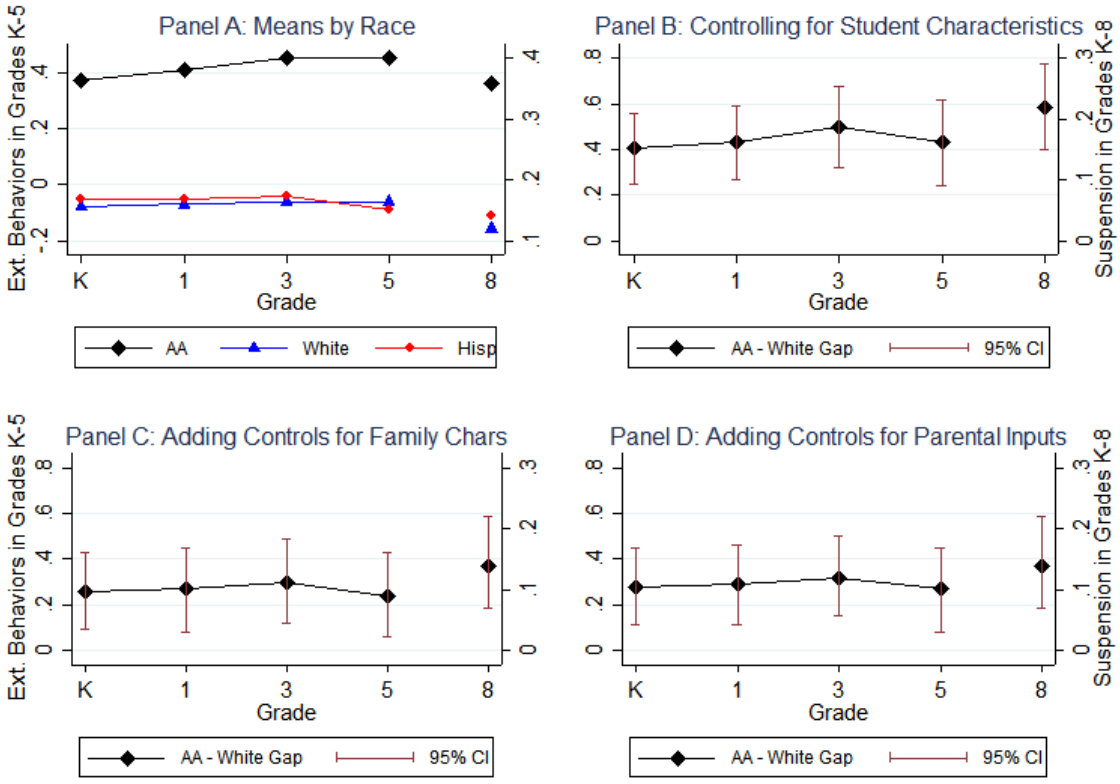

Figure 1.1 – Externalizing Behavior and School Suspension by Race

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:1999 panel weights.

Chapter 2

A Kindergarten Teacher Like Me:

The Role of Student-Teacher Race in Social-Emotional Development[37]

(with Michael Gottfried and Vi-Nhuan Le)

[37] Accepted manuscript at the American Educational Research Journal.

2.1 Introduction

Although people of color represent 41% of elementary and secondary students in the U.S., this proportion drops to only 16.5% when looking at the population of elementary and secondary teachers (Ingersoll & May, 2011). Even within urban and high-poverty schools, which enroll a high proportion of students of color and where teachers of color might be expected to be employed, White teachers comprise 71% and 65% of the faculty, respectively (Ingersoll & May, 2011). As the nation's student population is becoming increasingly diverse, policymakers have raised concerns that the teaching force will fall even further behind at reflecting the demographics of the student population it serves. To address this discrepancy between the racial and cultural backgrounds of students and their teachers, a number of initiatives have been designed to increase representation of teachers of color within teacher recruitment and preparation programs. For example, nearly half of U.S. states have implemented teacher recruitment programs or policies that focus on attracting teachers of color (Achinstein et al., 2010).

Underlying these calls for reforming teacher recruitment and retention are persistent disparities in school readiness between students of color and White students. Although most existing literature has focused on achievement differences (Fryer & Levitt, 2013), research has also reported racial/ethnic differences in teachers' ratings of social-emotional skills for students of color. As operationalized in this study, social-emotional skills relate to children's ability to manage their emotions (e.g., self-control) and behavior (e.g., externalizing behaviors), establish warm relationships with others (e.g., interpersonal skills), and display a positive attitude towards school (e.g., approaches to learning). García (2015) found that African-American kindergarteners were rated by their teachers as having poorer self-control,

worse approaches to learning, and higher frequencies of externalizing behaviors compared to White kindergarteners. In a similar vein, Latino children were rated by their kindergarten teachers as having poorer attention skills (Duncan & Magnuson, 2011) and lower persistence on classroom tasks than White children (West, Denton, & Reaney, 2001). From a policy standpoint, it is important to address these differences because studies have shown that social-emotional ratings upon kindergarten entry are predictive of a range of future outcomes, including achievement in later grades (Chetty et al., 2011), educational attainment (Lleras, 2008), and future income (Cunha & Heckman, 2009; Hall & Farkas, 2011). This study examines whether students of color have higher social-emotional ratings when they have a kindergarten teacher of the same race/ethnicity, and whether the findings vary by student or teacher characteristics.

2.2 Cultural Synchronicity as a Framework for Understanding Racial/Ethnic Differences in Social-Emotional Ratings

Some educators have argued that racial/ethnic differences in teachers' ratings of social-emotional skills for students of color can be addressed through an increase in the diversity of the teacher workforce, which would allow for greater rates of student-teacher race matching for students of color. Proponents of student-teacher race matching contend that teachers of color whose race/ethnicity matches their students may be more effective than White teachers because of cultural synchronicity (Irvine, 1988), in which teachers of color are able to capitalize on a shared cultural background. Through interactions with their parents and others within their community, students of color learn to adopt social behaviors, linguistic conventions, and cognitive proficiencies that reflect the norms of their culture

(Sameroff & Fiese, 2000). However, the schooling system, which is staffed predominantly by White teachers and administrators, tends to espouse White, middle-class standards of classroom deportment and behaviors (Morris, 2005), and these cultural discontinuities may lead students of color to struggle to conform to the academic and behavioral norms of school (Delpit, 1995). Cultural synchronization can enhance the educational outcomes of students of color because teachers of color whose race/ethnicity matches their students can leverage their common cultural experiences to create a culturally relevant learning environment and engage in culturally appropriate pedagogical strategies (Achinstein & Ogawa, 2011; Ladson-Billings, 1992).

Some researchers have argued that broad racial labels such as African American or Latino obscure the fact that members of the same racial group can have vastly different background experiences, cultural affiliations, language, and social class (Perez & Hirschman, 2010). Latinos, for example, encompass diverse groups such as Mexicans, Puerto Ricans, Cubans, and Central Americans (Etzioni, 2002), each of whom have unique assimilation histories and cultures. However, due to difficulties capturing cultural, economic, and social diversity within the same racial categories (Logan & Turner, 2013), the literature continues to proxy cultural synchronicity by determining whether there is matching between teachers' and students' race/ethnicity, broadly defined (Ingersoll & May, 2011).

To date, the notion of cultural synchronicity has been used primarily as a framework for understanding the academic achievement of students of color (Villegas & Irvine, 2010), but cultural synchronicity can also be thought of as a way to understand students' motivations and behaviors. Many White teachers may fail to recognize the behavioral

strengths of their students of color because those behaviors reflect the expected norms of another culture (Galindo & Fuller, 2010). This may result in students of color feeling marginalized or alienated because they do not perceive their schooling experiences as culturally relevant (Weinstein, Curran, & Tomlinson-Clarke, 2003). These types of cultural incompatibilities may strain student-teacher relationships (Crosnoe, Johnson, & Elder, 2004), which can result in impaired social-emotional outcomes and more frequent manifestations of disruptive behaviors (Monroe, 2005).

2.2.1 Studies Suggesting a Lack of Cultural Synchronicity within Schools

One consequence of the lack of cultural synchronicity is that White teachers' assessments of students of color may be less positive than those for White students. Studies have found this to be true for African American, Latino, and Native American students, but not for Asian American students (Chang & Demyan, 2007; McGrady & Reynolds, 2013). For this study, we focus on African American and Latino students because there is evidence that teachers direct fewer positive and neutral speech towards African American and Latino students (Tenenbaum & Ruck, 2007), hold lower academic expectations for them (Oates, 2003), and perceive them to be less mature than White students (Alexander, Entwisle, & Thompson, 1987). Studies have also found that teachers rate African American and Latino students as having lower levels of attentiveness (McGrady & Reynolds, 2013), lower task orientation (Sbarra & Pianta, 2001), and poorer work habits than White students (McClelland, Morrison, & Holmes, 2000).[38]

[38] Although these studies do not specifically report whether it was White teachers or teachers of color assigning the student ratings, the majority of teachers in these studies were White.

Of particular interest is whether assessments of externalizing behaviors, such as acting out, arguing, and getting angry, are particularly sensitive to the effects of having a teacher with a shared cultural background. Studies have shown that teachers assign students of color significantly lower ratings than White students in frustration tolerance (Sbarra & Pianta, 2001) and significantly higher ratings on aggressiveness and disobedience (Chang & Demyan, 2007). African American and Latino students also have considerably higher rates of suspensions and expulsions than do White students, with 15% of African American students and 6.8% of Latino students being suspended or expelled compared to 4.8% of White students (Hoffman, 2014). In addition, over 70% of the students involved in school-related arrests and referrals to law enforcement were African American or Latino (Rudd, 2014). Even after controlling for socioeconomic status, students of color continue to be overrepresented among expulsions and suspensions (McCray et al., 2015; Taylor, Crego, & Lane, 2014).

In light of research showing that the racial/ethnic gaps in expulsions and suspensions are not explained by more frequent or more serious misbehaviors by students of color (Skiba & Williams, 2014) and that students of color are punished more severely than White students for the same offense (Rudd, 2014), some educators have questioned whether these findings can be traced to cultural misunderstandings that lead teachers to interpret African American and Latino students' behaviors as "disruptive" (Noguera, 2008). For example, if a student of color is accustomed to responding to strong directive speech from his/her parents, then s/he may be considered noncompliant when responding to an indirect command from a White teacher (Delpit, 1995). 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 or smoking), whereas students of color are more likely to be referred for behaviors requiring subjective evaluations, such as defiance, excessive noise, or disrespectfulness (Gregory & Weinstein, 2008; Skiba et al., 2014).

2.2.2 Racial/Ethnic Congruence and Teachers' Assessments of Students' Social-Emotional Outcomes

From the perspective of cultural synchronicity, teachers of color may be able to draw upon their understanding of the lack of congruence between children's home culture and school's mainstream culture to consider a wider range of behaviors to be acceptable (Rimm-Kaufman, Pianta, & Cox, 2000), or to address any misbehaviors in a manner that does not further escalate tensions (Monroe & Obidah, 2004). If cultural synchronicity does indeed confer benefits to students of color in terms of a shared cultural understanding, then we would expect to observe more favorable assessments of students of color by teachers of color, especially with respect to externalizing behaviors. The empirical literature is generally sparse in this area, and the existing studies have reported mixed results. Analyzing ECLS-K:1999, Downey and Pribesh (2004) found that African American students, on average, were rated as exhibiting more externalizing behaviors than White students, but when same-race matching was taken into account, African American teachers rated their African American students as having fewer behavioral problems than White teachers rating White students. Using NELS:88, Dee (2005) analyzed students in eighth grade to examine how teachers of different races evaluated the same-race student. He found that students who did not share the race of their teacher were more likely to be labeled as disruptive and

inattentive. By way of contrast, Bradshaw, Mitchell, O'Brennan, and Leaf (2010) examined data from 21 elementary schools and found that the racial match between students and teachers did not reduce the rate of disciplinary referrals among African American students.

Other relevant work includes empirical examinations of how race/ethnicity interactions affect teacher assessments of other social-emotional skills. Using ECLS-K:1999, Jennings and DiPrete (2010) analyzed data on kindergarten and first-grade students and found no effect for African American or Latino students of having a same-race teacher on a composite social/behavioral scale comprised of approaches to learning, self-control, and interpersonal skills. In contrast, using 10th grade data from NELS:88, Ehrenberg, Goldhaber, and Brewer (1995) examined teachers' ratings on a composite scale that included items about students' ability to work hard and students' chances of going on to college. They found that relative to White teachers, Latino and African American teachers rated students of their same race/ethnicity more favorably. These results were similar to that of McGrady and Reynolds (2013), who analyzed 10th grade data from the nationally-representative ELS:2002. They reported that African American teachers rated African American students' academic progress more positively than did White teachers, and that Latino teachers were more likely than White teachers to believe that Latino students related well to others.

2.3 Current Study

The purpose of this study was to examine whether students of color have higher social-emotional ratings when they have a kindergarten teacher whose racial/ethnic group is the same as their own. Our study contributes to the literature on cultural synchronicity and

student-teacher race matching in two ways. First, we applied this framework to assess the most current nationally representative dataset available. National-level efforts to diversify the teacher workforce merit exploration with national-level data in order to make relevant conclusions. Although some studies have previously examined the impact of student-teacher race matching using large-scale data (e.g., Dee, 2005; Downey & Pribesh, 2004; Ehrenberg et al., 1995; Jennings & Diprete, 2010; McGrady & Reynolds, 2013), these datasets were collected more than a decade ago, and preceded many critical schooling policies and social trends that have shifted classroom curriculum and demographics. For example, these datasets preceded the implementation of No Child Left Behind, which increased the prominence of standards and accountability in K-12 schools nationwide, and raised concerns about whether the heightened focus on academics may negatively impact the attention span, self-regulation, and interpersonal skills of students (Stipek, 2006). In addition, these datasets predated the marked increases in the Latino population, which has accounted for over half the growth in the total U.S. population between 2000 and 2010 (Humes, Jones, & Ramirez, 2011). Finally, none of these prior studies that have used national-level data, explored these issues in the framework of cultural synchronicity as we have done.

Second, our study focuses on kindergarten students. We focused on kindergarten because studies have shown that teachers' assessments of children during kindergarten may influence their behaviors and outcomes in later years (Meisels, Steele, & Quinn-Leering, 1993). In a seminal study, Rist (2000) found that African American children who did not conform to White middle class norms in terms of appearance, behavior, or family structure were judged by their kindergarten teachers to be less likely to be successful at school, and received less teaching time, attention, and reward-directed behavior. Furthermore, realizing

the self-fulfilling prophecy of their teachers' low expectations, children who had been judged to be less likely to be successful at kindergarten demonstrated lower academic achievement and more disruptive behaviors at second grade. Thus, it is important to determine whether there is evidence of cultural incongruities early on in a student's educational career that can affect teachers' expectations and assessments of students, and to put policies in place that foster and support an educational environment that sets all children on a trajectory of school success.

2.4 Method

2.4.1 Sample

We analyzed data from the Early Childhood Longitudinal Study – Kindergarten Class of 2010-2011 (ECLS-K:2011). Created by the National Center for Education Statistics (NCES), ECLS-K:2011 is the most current data available for young children and their schooling experiences in early education. During the 2010-2011 academic year, NCES collected data about children in kindergarten through assessments and parent, teacher, and school administrator surveys.

Assessments and surveys were administered in the fall and spring of the kindergarten school year. ECLS-K:2011 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/ethnicity represented the third sampling unit. For this study, we refined our sample of student-level observations based on three primary criteria. First, we analyzed only African American and Latino teachers and students, with White teachers and students serving as the reference group. We did not

include Asian American, Native American, or Pacific Islander teachers and students in our study because there were very few same-race student-teacher matches for these groups (although our results are robust to their inclusion in the analyses). Second, in order to identify student and teacher race/ethnicity and include student controls in our regression specifications, we required non-missing data on relevant teacher and student background characteristics and at least one social-emotional assessment. Third, each student must have had at least one math or reading test score with a corresponding math or reading teacher assessment in both the fall and spring so that they could be included in the empirical test for teacher bias (available in the online appendix). These restrictions resulted in 9,140 students for the analytical sample. To comply with NCES reporting standards, sample sizes were rounded to the nearest ten.

2.4.2 Determination of Race/Ethnicity

Students' race/ethnicity was designated by NCES based on a combination of parent and school reports. Teachers' race/ethnicity was self-reported.

2.4.3 Social-Emotional Outcomes

In fall and spring of the kindergarten year, teachers completed survey assessments on those students in their classrooms who were in the ECLS-K:2011 sample. The surveys contained a series of question items that were then combined to create the social-emotional scales. In this study, we analyzed all teacher-rated social-emotional scales in the dataset. These social-emotional scales were constructed based on the Social Skills and Rating System (SSRS; Gresham & Elliott, 1990). NCES adapted the SSRS to create its own teacher

Social Rating Scales (SRS). The SRS in the dataset consisted of five scales that gauged the extent that students exhibited various social-emotional skills. The 5-item externalizing behaviors scale assessed the frequency by which a child argued, fought, got angry, acted impulsively, and disturbed ongoing activities. The 4-item internalizing behaviors scale measured the extent that the child exhibited anxiety, loneliness, low self-esteem, and sadness. The 5-item interpersonal skills scale measured the frequency by which a child got along with others, formed and maintained friendships, helped other children, showed sensitivity to the feelings of others, and expressed feelings, ideas, and opinions in positive ways. The 7-item approaches to learning scale rated the frequency that the child kept his or her belongings organized, showed eagerness to learn new things, adapted to change, persisted in completing tasks, paid attention, and followed classroom rules. The 4-item self-control scale measured the extent that the child was able to control his or her temper, respected others' property, accepted his or her peers' ideas, and handled peer pressure.

Teachers' responses to each 4-point Likert item were averaged to create scale scores. Note, however, that individual question items were not available, even in the restricted user's manual of the dataset. Higher scale scores denoted more frequently exhibited behaviors. For the externalizing and internalizing scales, a higher score reflects a less favorable outcome. On the three other scales, a higher score reflects a more favorable outcome. All the scales had high internal consistency, with the alpha reliability coefficients ranging from 0.79 to 0.89 (Tourangeau et al., 2012). We limited our sample so that the same teacher assessed each child in the fall and spring surveys, thereby providing for consistency between assessment waves.

2.4.4 Control Variables

Our analysis included control measures for a commonly-upheld set of demographics associated with differences in social-emotional ratings, including gender, age, race, and an indicator for English language learner (ELL) based on the primary language spoken at home (Duncan & Magnuson, 2011; Garcia, 2015). We also included a parental-rated scale of child health (1 being highest, 5 being lowest) because poor health has been linked to weaker social-emotional functioning (Meijer et al., 2000).

Research supports the link between childcare in the years prior to kindergarten with social-emotional outcomes in kindergarten (Loeb et al., 2007; Magnuson et al., 2007; Yamauchi & Leigh, 2011), although the literature is mixed with respect to the direction of this relationship. Nonetheless, we included the following commonly employed childcare measures as indicators for care in the year prior to kindergarten: center-based care, Head Start, non-relative care, and relative care (parental care serves as the reference group). We also included an indicator for whether the child has ever participated in any type of center-based care as well as the number of non-parental weekly care hours to control for the fact that intensity of care might be linked to social-emotional outcomes in kindergarten (Loeb et al., 2007).

Family characteristics included socioeconomic status (SES), family structure, and home environment. Research has linked lower SES to worsened social-emotional functioning in children (Bradley & Corwyn, 2002). Therefore, we included a NCES-created composite of SES, which was comprised of items relating to parental educational attainment, occupational prestige, and family income. We separated this composite into quartile indicators for our regression analysis. We also included information about whether

two biological or adoptive parents were present in the student's home, as parental structure can influence children's social-emotional development (Amato, 2005; Kupersmidt et al., 1995).

In addition, we replicated two scales from Votruba-Drzal et al.'s (2008) study that assessed the quality of the home environment. We controlled for home environment because previous studies have found the quality of the home environment to be positively correlated with students' social-emotional outcomes (Harden & Whittaker, 2011). The first scale consisted of 10 items measured on a 4-point Likert metric. The scale assessed the frequency with which parents engaged the child in activities that promoted cultural, academic, or social enrichment. These activities included playing games, singing songs, and reading books. The second scale, comprised of 15 dichotomously-scored items, assessed children's access to learning materials. This scale assessed whether in the past month the child engaged in activities such as visiting a book store, taking music lessons, or attending tutoring lessons. These composite scores were transformed to have a mean of zero and standard deviation of one within the sample.

We also included a key classroom characteristic that varied by student: the percentage of same-race peers in the classroom. Previous literature has found a positive association between having more peers of the same race/ethnicity and positive socio-emotional outcomes (Benner & Crosnoe, 2011). Because the proportion of same-race classmates was highly correlated with having a same-race teacher for students of color in our sample ($\rho = 0.55$), it was imperative to control for this percentage to separately identify the association between having a same-race teacher and student outcomes.

2.4.5 Analytic Approach

As conceptualized in Figure 2.1, social-emotional outcomes were influenced by cultural synchronicity, as represented by the teacher-student race matching measure, as well as other key child-level and classroom control variables. Mapping from our conceptual model to an analytic model that would allow us to empirically assess whether students of color demonstrated social-emotional gains from having a teacher of color of the same race/ethnicity, we used a gain-score model of student-level social-emotional outcomes. We adopted a similar approach as other studies that have examined student-teacher race matching (Fairlie et al., 2014) and estimated the gain-score (spring rating minus fall rating), y_{ic}, as:

(2.1) $y_{ic} = \alpha_0 + \alpha_1 PCteach_c + \alpha_2 PCstud_i + \delta PCteach_c * PCstud_i + X_i'\beta + \varepsilon_{ic}$

where students were indexed by *i* and classrooms by *c*. In the model, the variable $PCteach_c$ equaled one if teacher *c* was African American or Latino. For now, teachers were included in an aggregate person of color category although we later disaggregated this category by specific race. Analogously, $PCstud_i$ equaled one if student *i* was a person of color – African American or Latino. The term $\boldsymbol{X_i}$ in the equation represents all control variables described above as well as a squared term for age, which allowed us to estimate non-linear effects of age.

Of primary interest was the parameter δ, which was the interaction between being a student of color and having a teacher of color. This interaction term measured whether students of color had larger gains (y_{ic}) relative to White students when each had a teacher of color compared to this difference in gains (y_{ic}) when each was taught by a White teacher. In other words, δ measured the differential gains in social-emotional outcomes between

students of color and White students based on having a teacher of color versus having a White teacher.

In our preferred specification, we included classroom fixed effects. That is, we included indicator variables for each classroom with one classroom omitted to serve as the reference category. Including indicators for classrooms means that we controlled for classroom-to-classroom (and hence teacher-to-teacher) differences in the sample. Importantly, students in the kindergarten classes in ECLS-K were contained in one classroom, and they did not move to other classrooms throughout the day. Thus, by including classroom fixed effects we have held constant all unobserved classroom-level experiences and teacher influences during the school year. Controlling for classroom-specific factors alleviates issues such as the possibility that different instructors systematically assessed their students differently. Note that when holding classroom factors constant, any classroom-level variables would drop away as they would be collinear with the indicator for classroom. This means that we dropped the teacher of color indicator variable as it was collinear with this classroom indicator. We also allowed the person of color student-teacher match to vary by whether the match was of the same race/ethnicity or different races/ethnicities. This model with classroom fixed effects, γ_c, can be represented as:

(2.2) $$y_{ic} = \alpha_0 + \alpha_2 PCstud_i + \delta_1 PCteach_c * PCstud_i * same_{ic} + \delta_2 PCteach_c * PCstud_i * diff_{ic} + \gamma_c + X_i'\beta + \varepsilon_{ic}$$

where the indicator $same_{ic}$ equaled one if student *i* and classroom teacher *c* shared the same race/ethnicity and the indicator $diff_{ic}$ equaled one if student *i* and classroom teacher *c* did not share the same race/ethnicity.

By holding constant all classroom-specific factors with the classroom fixed effect, the key source of variation used to identify the effects of student-teacher race matching for students of color (the coefficients δ_1 and δ_2) occurred within each classroom rather than across classrooms. As the changes in social-emotional outcomes of students of color was measured relative to changes of White students within a classroom, only classrooms with at least one student of color and one White student contributed to the estimation of δ_1 and δ_2.

Finally, in all models, standard errors were clustered at the classroom level. Because students in the same classroom shared common but unobservable characteristics and experiences, clustering student data at the classroom level provided for a corrected error term given this non-independence of student observations. In addition, observations were weighted using ECLS-K:2011 sample weights (W12T0) to estimate representative effects.

Given our preferred specification, there were two primary threats to identifying the relationship between student-teacher race match and student outcomes: nonrandom sorting of students to classrooms and race-based teacher bias in teacher-assessed outcomes. Our tests for sorting and bias, which can be found in Appendix A.2, suggest that neither of these issues were significant factors in our analyses.

2.5 Results

2.5.1 Descriptive Statistics

Table 2.1 presents descriptive statistics for the outcomes used in the main empirical analysis. Panel A of Table 2.1 gives student and teacher composition by race/ethnicity and Panel B gives the mean and standard deviation of social-emotional outcomes by race/ethnicity. Approximately 14% of all students were students of color matched with

teachers of color; within this subsample, 12% were same-race matches and 2% were matches of different races.

2.5.2 Social-Emotional Outcomes

We report estimates of the same-race and different-race interaction variables for the five social-emotional gain-score outcomes in Table 2.2. Column (3) gives estimates from the preferred specification described in equation (2.2). Classroom fixed effects and student controls were omitted from the regressions in Column (1). The importance of controlling for unobserved classroom influences is evident in column (2) where classroom fixed effects were added; the magnitudes of the point estimates change drastically moving from column (1) to column (2). Adding student controls in column (3), on the other hand, did not substantially affect the estimates or their precision. The same-race coefficient was null for internalizing behaviors, interpersonal skills, approaches to learning, and self-control. However, the same-race coefficient for externalizing behaviors was significantly different from zero, and indicates that the student of color versus White gap in disruptive behavior improved by 0.26 standard deviations for students of color matched with a same-race teacher. We found no significant effect of the different-race person-of-color interaction on student outcomes in our preferred model.

2.5.3 Externalizing Behaviors

Because cultural synchronicity may be most demonstrably manifested on teachers' assessments of students' externalizing behaviors, we focused the rest of our analysis on externalizing behaviors. In the previous analysis, we assumed the coefficient associated with

same-race interaction between African American teachers and students was equivalent to the coefficient associated with same-race interaction between Latino teachers and students. To further analyze specific teacher-student race/ethnicity interactions, the person of color student-teacher indicator was broken down into four categories: Latino student-Latino teacher; African American student-Latino teacher; Latino student-African American teacher; African American student-African American teacher. The results of this analysis are presented in Table 2.3 for externalizing behaviors. The associated p-value for a t-test for the equality of the Latino-Latino and African American-African American interaction effect is also presented in Table 2.3. Both same-race interaction coefficients were significantly different from zero. Although the African American same-race effect was nearly double the magnitude of the Latino same-race effect, we could not reject the equality of these two coefficients. Additionally, we found no statistically significant effect of either different-race interaction.

To examine what might be driving the relationship between person of color teacher-student race match and externalizing behaviors, we broke down the same-race interaction by student and family traits in Table 2.4. The first column presents results for the aggregated same-race interaction term for teachers and students of color and the second and third columns describe the African American student-teacher match and the Latino student-teacher match, respectively. For comparison, we provide the baseline estimates from column (3) of Table 2.2 and Table 2.3 in the first row.

As Latino teachers may be able to better connect with Latino ELL students working to overcome a language barrier at school, we separated the same-race interaction for Latino students by ELL status. We restricted our analysis to Latino ELLs because they comprised

over 99% of all ELLs with a same-race teacher. On average, Latino ELL students exhibited a relatively large, favorable response to Latino teachers; race-matched ELL students experienced an improvement in externalizing behaviors of 0.40 standard deviations whereas the effect on non-ELL students was not statistically different from zero. Furthermore, the null hypothesis that the ELL and non-ELL same-race interactions were equal was rejected at the 5% level.

Thus far we have used sharing the same race/ethnicity exclusively as our proxy for cultural synchronicity between teachers and students. Another way that teachers may form connections with students, irrespective of race/ethnicity, is through language. We explored this notion by creating a new category of Spanish-speaking African American and White teachers and examined the interaction between these Spanish-speaking teachers and Latino ELL students. Interestingly, we found no evidence that Latino ELL students with non-Latino, Spanish-speaking teachers improved their externalizing behavior ratings. We also found no significant difference in the same-race interaction by gender, SES (students are designated as either high or low SES based on whether their SES composite score was above or below the SES median for students of color in a classroom with a teacher of the same race/ethnicity), or family structure.

2.6 Discussion

As students of color continue to grow as a percentage of the U.S. population, many policymakers have called for increased representation of people of color among the teacher workforce in order to reflect these changing student demographics (Ingersoll & May, 2011). To date, much of the empirical research about cultural synchronicity and student-teacher

race matching has focused on achievement outcomes (Villegas & Irvine, 2010). Our study is one of the first empirical studies to examine students' social-emotional outcomes from a cultural synchronicity perspective while simultaneously employing large-scale national data as well as tests for teacher bias and student selection into classrooms. Analyzing the most current national-level dataset of U.S. kindergarteners, we found that same-race matching was unrelated to teachers' ratings of children's internalizing problem behaviors, interpersonal skills, approaches to learning, or self-control. This is consistent with the findings from earlier studies (Jennings & DiPrete, 2010). The mostly null findings may reflect the limitations of our analysis, where we could not adequately capture variation in cultural experiences for members of the same racial/ethnic group, so we had to proxy cultural synchronization through broad categorizations of race/ethnicity matching.

That being said, having a student-teacher race match for students of color was associated with a decline in externalizing behaviors by 0.26 standard deviations. This is comparable to the initial (fall) regression-adjusted gap between students of color and White students in disruptive behavior of 0.27 standard deviations, which suggests that having a student-teacher race match for students of color nearly extinguished the externalizing behavior gap by the end of kindergarten. Furthermore, using Cohen's (1988) guidelines, where a regression coefficient of 0.15 is considered a medium effect and a regression coefficient of 0.35 is considered a large effect, the effect sizes for student-teacher race match is considered medium for Latinos and large for African Americans.

Notably, Latino ELL students demonstrated much lower externalizing behaviors when paired with a Latino teacher, but there was no improvement in externalizing behaviors when Latino ELL students were paired with a Spanish-speaking, non-Latino teacher. This

finding bolsters the argument that cultural synchronicity and a shared cultural identity and background, as opposed to an ability to communicate using a common language, is key to reducing externalizing behaviors.

Given these findings, there are three important implications from our study. First, the results suggest that increasing the representation of teachers of color may be a promising way to address externalizing behaviors. Policies regarding disciplinary actions have come under scrutiny in light of studies showing that students of color are suspended or expelled at three times the rates of White students (U.S. Department of Education for Civil Rights, 2014). In schools that reported expulsions under zero tolerance policies, Latino and African American students comprised 45% of the student population, but represented 56% of the students expelled under such policies (U.S. Department of Education for Civil Rights, 2014). Studies have shown that many of the disciplinary actions are for minor misbehaviors such as insubordination, which may reflect subjective or biased interpretations on the parts of teachers (American Psychological Association Zero Tolerance Task Force, 2008). Our study provides some support for this notion, as students of color were less likely to be perceived as acting out when assessed by their same-race teachers. This finding suggests that the disparate impact of disciplinary policies on students of color could be alleviated by hiring teachers who share the same cultural background and experiences as their students, and who may interpret externalizing behaviors in the context of students' cultural backgrounds.

Second, this study found that student-teacher race matching plays a significant role in student behavior at the onset of schooling. Therefore, elementary schools can use this information to help design environments that are most supportive for student behavior given

a set of student populations and teaching resources. In light of the fact that same-race matching could almost erase the gap in externalizing behaviors by the end of the academic year, this study urges interventions be implemented early in the school year, thereby strengthening the potential for school success from the very start of formal education.

Finally, the results suggest increased training for all teachers to become more familiar with the diverse cultural backgrounds of students of color. Many of the behavioral management strategies found in teacher education and professional development programs reflect mainstream values, and may not adequately capture culturally responsive strategies that reflect the speech patterns, voice tones, word choices, and facial expressions familiar to students of color (Monroe & Obidah, 2004). In a review of school-wide positive behavior supports that emphasized culturally responsive practices, Fallon, O'Keeffe, and Sugai (2012) recommended several best practices that may increase teachers' self-awareness of their own biases that could alleviate unduly harsh assessments of students' externalizing behaviors. These strategies include identifying the characteristics of students who are most likely to be punished, encouraging students to share their home culture and learning histories so that teachers may better understand students' motivations for their behaviors, and refraining from punishing students for behaviors that may be appropriate in other contexts. As policymakers and educators grapple with how to mitigate the disproportionate racial/ethnic patterns in suspension and disciplinary actions, adopting diverse behavioral management strategies that reflect the cultural backgrounds of students of color may prove to be a particularly promising strategy, although more research is warranted to understand whether different approaches are needed for Latinos and African Americans, given the different cultural histories of experience for members of these racial/ethnic groups.

In sum, the findings from our study provide tangible policy and practice strategies to address issues of culture, race/ethnicity, and equity in developing supportive classroom environments as the demographic makeup of our nation's students continues to diversify. The approach taken in this study also provides implications for developing the future direction of educational research. Prior research has often existed in silos – both in terms of discipline and methodological approach. This study, which involved multiple disciplines in the conceptualization of the classroom vis-à-vis culture and race/ethnicity, exemplifies the necessity of future research to develop a multifaceted, interdisciplinary perspective. Further, the theory was not disconnected from the empirical model, as is often the case. Rather, both prior qualitative and quantitative research was used to directly build not only our specification, but also to identify appropriate measures for the model.

Future educational research must not simply be about one discipline "informing" the other. Instead, the approach taken by educational researchers should be to capitalize on how each field intersects such that we, as one field, can conceptualize and measure the nuances of the classroom in ways that simultaneously advance research and ensure child success. In this vein, our study calls on the next generation of data collection efforts to draw upon multiple disciplines to more effectively capture these nuances. For instance, the national data used in our study takes into account much of the context of the classroom, but is limited by lacking detailed information on teachers' background experiences, cultural affiliations, language, and social class. While we have these data for children, the perspective remains imbalanced without analogous teacher data. Current national datasets, like the one in this study, certainly help to provide key evidence of trends and patterns to build supportive learning systems for children. However, to move the field forward over the next century, we

call for new data to rely on the interdisciplinary and multi-faceted body of research to move research and analysis forward rather than rigidly mirroring previous data collection efforts. Doing so not only makes for stronger research, but also more appropriately reflects the changing demographics of our nation.

References (Chapter 2)

Achinstein, B. & Ogawa, R.T. (2011). *Change(d) agents: New teachers of color in urban schools.* New York: NY: Teachers College Press.

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.

Amato, P.R. (2005). The impact of family formation change on the cognitive, social, and emotional well-being of the next generation. *The Future of Children, 15*(2), 75-96.

Alexander, K., Entwisle, D., & Thomson, M. (1987). School performance, status relations, and the structure of sentiment-bringing the teacher back in. *American Sociological Review, 52*(5), 665-682.

American Psychological Association Zero Tolerance Task Force (2008). *Are zero tolerance policies effective in the schools? An evidentiary review and recommendations*. Washington, DC: Author.

Benner, A.D., & Crosnoe, R. (2011). The racial/ethnic composition of elementary schools and young children's academic and socioemotional functioning. *American Educational Research Journal*, *48*(3), 621-646.

Bradley, R.H., & Corwyn, R.F. (2002). Socioeconomic status and child development. *Annual Review of Psychology, 53,* 371-399.

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.

Chang, D. F. & Demyan, A. (2007). Teachers' stereotypes of Asian, Black, and White students. *School Psychology Quarterly, 22*(2), 91-114.

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.

Crosnoe, R., Johnson, M.K. & Elder, G.H. (2004). Intergenerational bonding in school: The behavioral and contextual correlates of student-teacher relationships. *Sociology of Education, 77*(1), 60-81.

Cohen, J. (1988). *Statistical power analysis for the behavior sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cunha, F. & Heckman, J. (2009). The economics and psychology of inequality and human development. *Journal of the European Economic Association,* 7(2-3), 320-364.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review*, *95(*2*),* 158–65.

Delpit, L. (1995). *Other people's children: Cultural conflict in the classroom*. New York: New Press.

Downey, D., & Pribesh, S. (2004). When race matters: Teachers' evaluations of students' classroom behavior. *Sociology of Education, 77*(4), 267-282.

Duncan, G.J., & Magnuson, K. (2011). The nature of early achievement skills, attention, and behavior problems. In G.J. Duncan and R.J. Murnane (eds). *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances* (pp. 47-69). New York: Russell Sage.

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.

Etzioni, A. (2002). *Inventing Hispanics*: *A diverse minority resists being labeled*. Washington, DC: Brookings Institution.

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.

Fallon, L.M., O'Keeffe, B., & Sugai, G. (2012). Consideration of culture and context in school-wide positive behavior support: A review of current literature. *Journal of Positive Behavioral Interventions, 14*(4), 209-219.

Fryer R.G., & Levitt, S.D. (2013). Testing for racial differences in the mental ability of young children. *The American Economic Review*, *103*(2), 981-1005.

Galindo, C., & Fuller, B. (2010). The social competence of Latino kindergartners and growth in mathematical understanding. *Developmental Psychology*, *46*(3), 579-592.

García, E. (2015). *Inequalities at the starting gate: Cognitive and noncognitive skills gaps between 2010-2011 kindergarten classmates.* Washington, DC: Economic Policy Institute.

Gregory, A., & Weinstein, R.S. (2008). The discipline gap and African Americans: Defiance of cooperation in the high school classroom. *Journal of School Psychology*, *46*(4), 455-475.

Gresham, F.M., & Elliot, S.N. (1990). *Social skills rating system (SSRS)*. Circle Pines, MN: American Guidance Service. Hall, M.Y., & Farkas, G. (2011). Adolescent cognitive skills, attitudinal/behavioral traits and career wages. *Social Forces 89*(4), 1261-1285.

Harden, B.J., & Whittaker, J.E.V. (2011). The early home environment and developmental outcomes for young children in the child welfare system. *Children and Youth Services Review, 33*(8), 1392-1403.

Hoffman, S. (2014). Zero benefit: Estimating the effect of zero tolerance discipline policies on racial disparities in school discipline. *Educational Policy, 28*(1), 69-95.

Humes, K.R., Jones, N.A., & Ramirez, R.R. (2011). *Overview of race and Hispanic origin*: *2010*. Washington, DC: U.S. Census Bureau.

Ingersoll, R.M., & May, H. (2011). Recruitment, retention, and the minority teacher shortage. Philadelphia, PA: Consortium for Policy Research in Education.

Irvine, J.J. (1988). An analysis of the problem of the disappearing Black educator. *Elementary School Journal, 88*(5), 503-514.

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.

Kupersmidt, J.B., Griesler, P.C., DeRosier, M.E., Patterson, C.J., & Davis, P.W. (1995). Childhood aggression and peer relations in the context of family and neighborhood factors. *Child Development*, *66*(2), 360-375.

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

Lleras, C. (2008). Do skills and behaviors in high school matter? The contribution of noncognitive factors in explaining differences in educational attainment and earnings. *Social Science Research, 37*(3), 888-90.

Loeb, S., Bridges, M., Bassok, D., Fuller, B., & Rumberger, R.W. (2007). How much is too much? The influence of preschool centers on children's social and cognitive development. *Economics of Education Review, 26,* 52-66.

Logan, J.R., & Turner, R.N. (2013). *Hispanics in the U.S.: Not only Mexicans*. Providence, RI: Brown University.

Magnuson, K.A., Ruhm, C., & Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance? *Economics of Education Review, 26,* 33-51.

McClelland, M.M., Morrison, F.J., & Holmes, D.L. (2000). Children at risk for early academic problems: the role of learning-related social skills. *Early Childhood Research Quarterly*, *15*(3), 307-329.

McCray, C. R., Beachum, F. D., & Yawn, C. D. (2015). Saving our future by reducing suspensions and expulsions among African American males. *Journal of School Leadership*, *25*(2), 345.

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

Meijer, S.A., Sinnema, G., Bijstra, J.O., Mellenbergh, G.J., & Wolters, W.H.G. (2000). Social functioning in children with a chronic illness. *Journal of Child Psychology and Psychiatry*, *41*(3), 309-317.

Meisels, S.J., Steele, D.M., & Quinn-Leering, K. (1993). Testing, tracking, and retaining young children: An analysis of research and social policy. In B. Spodek (Ed.), *Handbook of research on the education of young children* (pp. 279–292). New York: Macmillan Publishing Company.

Monroe, C.R. (2005). Understanding the discipline gap through a cultural lens: Implications for the education of African American students. *Intercultural Education, 16*, 317–330.

Monroe, C.R., & Obidah, J.E. (2004). The influence of cultural synchronization on a teacher's perceptions of disruption: A case study of an African American middle-school classroom. *Journal of Teacher Education*, *55*(3), 256-268.

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.

Noguera, P.A. (2008). What discipline is for: Connecting students to the benefits of learning. In M. Pollock (Ed.), Everyday antiracism: Getting real about race in schools. (pp. 132–137). New York: The New Press.

Oates, G.L. (2003). Student-teacher racial congruence, teacher perceptions, and student performance. *Social Science Quarterly*, *84*(3), 508-525.

Perez, A.D., & Hirschman C. (2009). The changing racial and ethnic composition of the U.S. population: Emerging American identities. *Population Development Review*, *35*(1), 1-51.

Rimm-Kaufman, S.E., Pianta, R.C., & Cox, M.J. (2000). Teachers' judgments of problems in the transition to kindergarten. *Early Childhood Research Quarterly*, *15*(2), 147-166.

Rist, R. (2000). Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education. *Harvard Educational Review, 70*(3)¸257-301.

Rudd, T. (2014). *Racial disproportionality in school discipline.* Columbus, OH: Kirwan Institute.

Sameroff, A.J., & Fiese, B.H. (2000). Transactional regulation: The developmental ecology of early intervention. In S. Meisels & J. Shonkoff (Eds.), *Early intervention: A handbook of theory, practice, and analysis* (pp. 135–159). New York, NY: Cambridge University Press.

Sbarra, D.A., & Pianta, R.C. (2001). Teacher ratings of behavior among African American and Caucasian children during the first two years of school. *Psychology in the Schools*, *38*(3), 229-238.

Skiba, R.J., Chung, C., Trachok, M., Baker, T.L., Sheya, A., & Hughes, R.L. (2014). Parsing disciplinary disproportionality: Contributions of infraction, student, and school characteristics to out-of-school suspension and expulsion. *American Educational Research Journal, 51*(4), 640-670.

Skiba, R.J., & Williams, N. T. (2014). *Are black kids worse? Myths and facts about racial differences in behavior.* Bloomington, IN: Indiana University.

Stipek, D. (2006). No child left behind comes to preschool. *The Elementary School Journal, 106*(5), 455–466.

Taylor, J., Crego, M., & Lane, P. (2014). *Not measuring up*: *The state of school discipline in Massachusetts*. Boston, MA: Lawyers' Committee for Civil Rights and Economic Justice.

Tenenbaum, H.R., & Ruck, M.D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology, 99* (2), 253-273.

Tourangeau, K., Nord, C., Lê, T., Sorongon, A. G., Hagedorn, M.C., Daly, P., & Najarian, M. (2012). *Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K: 2011), User's Manual for the ECLS-K: 2011 Kindergarten Data File and Electronic Codebook* (NCES 2013-061). Washington, DC: US Department of Education, National Center for Education Statistics.

U.S. Department of Education for Civil Rights. (2014). *Data snapshot: School discipline*. Washington, DC: Author.

Villegas, A.M., & Irvine, J.J. (2010). Diversifying the teaching force: An examination of major arguments. *The Urban Review*, *42*(3), 175-192.

Votruba-Drzal, E., Li-Grining, C.P., & Maldonado-Carreño, C. (2008). A developmental perspective on full-versus part-day kindergarten and children's academic trajectories through fifth grade. *Child Development, 79*(4), 957-978.

Weinstein, C.S., Curran, M., & Tomlinson-Clarke, S. (2003). Culturally responsive classroom management: Awareness into action. *Theory into Practice, 42*(4), 269–276.

West, J., Denton, K., & Reaney, L. (2001). *The kindergarten year* (NCES 2001–023). Washington, DC: National Center for Education Statistic.

Yamauchi, C., & Leigh, A. (2011). Which children benefit from non-parental care? *Economics of Education Review, 30*, 1468-1490.

Tables (Chapter 2)

Table 2.1 – Descriptive Statistics

	Students			Teachers	
	Mean	Total obs.		Mean	Total obs.
Panel A. Student and teacher shares by race/ethnicity		9,140			2,420
White	0.61			0.82	
Latino	0.23			0.11	
African American	0.16			0.07	
Share of obs. person of color student–teacher match (same race)	0.12				
Share of obs. person of color student–teacher match (diff. race)	0.02				
	White		Latino	African American	
Panel B. Mean student outcomes by race/ethnicity					
Externalizing behaviors (fall)	-0.03		-0.04	0.17	
	(0.98)		(0.96)	(1.09)	
Externalizing behaviors (spring)	-0.04		-0.05	0.24	
	(0.98)		(0.95)	(1.11)	
Externalizing observations: 8,860					
Internalizing behaviors (fall)	0.01		-0.02	-0.01	
	(0.98)		(1.02)	(1.05)	
Internalizing behaviors (spring)	-0.02		0.02	0.07	
	(0.97)		(1.03)	(1.08)	
Internalizing observations: 8,810					
Interpersonal skills (fall)	0.06		-0.06	-0.09	
	(0.98)		(1.00)	(1.05)	
Interpersonal skills (spring)	0.07		0.01	-0.17	
	(0.99)		(0.98)	(1.04)	
Interpersonal observations: 8,480					
Approaches to learning (fall)	0.07		-0.08	-0.12	
	(0.99)		(1.00)	(1.02)	
Approaches to learning (spring)	0.06		-0.06	-0.16	
	(0.98)		(1.00)	(1.04)	
Approaches observations: 9,120					
Self-control (fall)	0.05		-0.03	-0.13	
	(0.98)		(1.00)	(1.05)	
Self-control (spring)	0.06		-0.01	-0.21	
	(0.98)		(0.99)	(1.07)	
Self-control observations: 8,380					

Notes. All student outcomes are standardized to be mean zero and standard deviation one by assessment wave.

Table 2.2 – Estimated Effects of Teacher-Student Race Matching

	(1)	(2)	(3)
Externalizing behaviors			
Persons of color student-teacher interaction:	0.00	-0.21**	-0.26**
same race	(0.05)	(0.09)	(0.10)
Persons of color student-teacher interaction:	-0.09	0.05	0.05
different race	(0.07)	(0.15)	(0.15)
Internalizing behaviors			
Persons of color student-teacher interaction:	-0.08	0.00	0.03
same race	(0.06)	(0.13)	(0.14)
Persons of color student-teacher interaction:	-0.17**	-0.02	-0.03
different race	(0.08)	(0.13)	(0.13)
Interpersonal skills			
Persons of color student-teacher interaction:	0.07	0.11	0.13
same race	(0.07)	(0.10)	(0.11)
Persons of color student-teacher interaction:	0.14	0.07	0.07
different race	(0.09)	(0.13)	(0.13)
Approaches to learning			
Persons of color student-teacher interaction:	0.01	0.04	0.07
same race	(0.06)	(0.09)	(0.10)
Persons of color student-teacher interaction:	0.15**	0.02	0.02
different race	(0.06)	(0.12)	(0.11)
Self-control			
Persons of color student-teacher interaction:	-0.03	-0.01	-0.01
same race	(0.07)	(0.12)	(0.13)
Persons of color student-teacher interaction:	0.16**	-0.06	-0.06
different race	(0.08)	(0.13)	(0.13)
Fixed effects			
Classroom	No	Yes	Yes
Controls			
Student controls	No	No	Yes

Notes. Each pair of same race and different race interactions represents a separate OLS regression.
*p < 0.1, **p < 0.05, ***p<0.01

Table 2.3 – Estimated Effects of Specific Teacher-Student Race Interactions on Externalizing Behaviors

	Teacher race/ethnicity	
	African American	Latino
Outcome: Externalizing behaviors		
Observations: 8,860		
Student race/ethnicity		
African American	**-0.40*** **	-0.27
	(0.15)	(0.20)
Latino	0.07	**-0.22***
	(0.17)	**(0.13)**
t-test: Equality of same-race/ethnicity effect (p-value)	0.35	

Notes. The set of four interactions is from the same OLS regression. This table displays the results from our preferred specification with classroom fixed effects and student controls.
*p < 0.1, **p < 0.05, ***p<0.01

Table 2.4 – Estimated Effects of Teacher-Student Race Matching on Externalizing Behaviors, by Student and Family Traits

	Combined African Americans and Latinos	African Americans	Latinos
Outcome: Externalizing behaviors			
Persons of color student-teacher interaction: same race	-0.26**	-0.40***	-0.22*
	(0.10)	(0.15)	(0.13)
ELL × same ethnicity	----	----	-0.40***
			(0.15)
Non-ELL × same ethnicity	----	----	-0.12
			(0.13)
t-test: Equality of the two effects (p-value)	----	----	0.04
ELL × same ethnicity	----	----	-0.45***
			(0.16)
ELL × Spanish-speaking non-Latino teacher	----	----	-0.15
			(0.12)
t-test: Equality of the two effects (p-value)	----	----	0.06
Male × same race/ethnicity	-0.23**	-0.37**	-0.19
	(0.11)	(0.16)	(0.14)
Female × same race/ethnicity	-0.30***	-0.42***	-0.26*
	(0.11)	(0.16)	(0.14)
t-test: Equality of the two effects (p-value)	0.36	0.68	0.54
High SES × same race/ethnicity	-0.23**	-0.37**	-0.19
	(0.11)	(0.16)	(0.13)
Low SES × same race/ethnicity	-0.33***	-0.39**	-0.36**
	(0.12)	(0.18)	(0.16)
t-test: Equality of the two effects (p-value)	0.34	0.89	0.18
Two biological parents × same race/ethnicity	-0.25**	-0.28	-0.22*
	(0.11)	(0.17)	(0.13)
Other family type × same race/ethnicity	-0.28**	-0.44***	-0.18
	(0.11)	(0.16)	(0.15)
t-test: Equality of the two effects (p-value)	0.72	0.26	0.72

Notes. Baseline estimates from column (3) of Table 2.2 and Table 2.3 are provided for comparison in the first row. In the first column, each pair of trait interactions represents a separate OLS regression. In the last two columns, both pairs of trait interactions for each race were estimated in the same regression (four coefficients altogether).
*p < 0.1, **p < 0.05, ***p<0.01

Figures (Chapter 2)

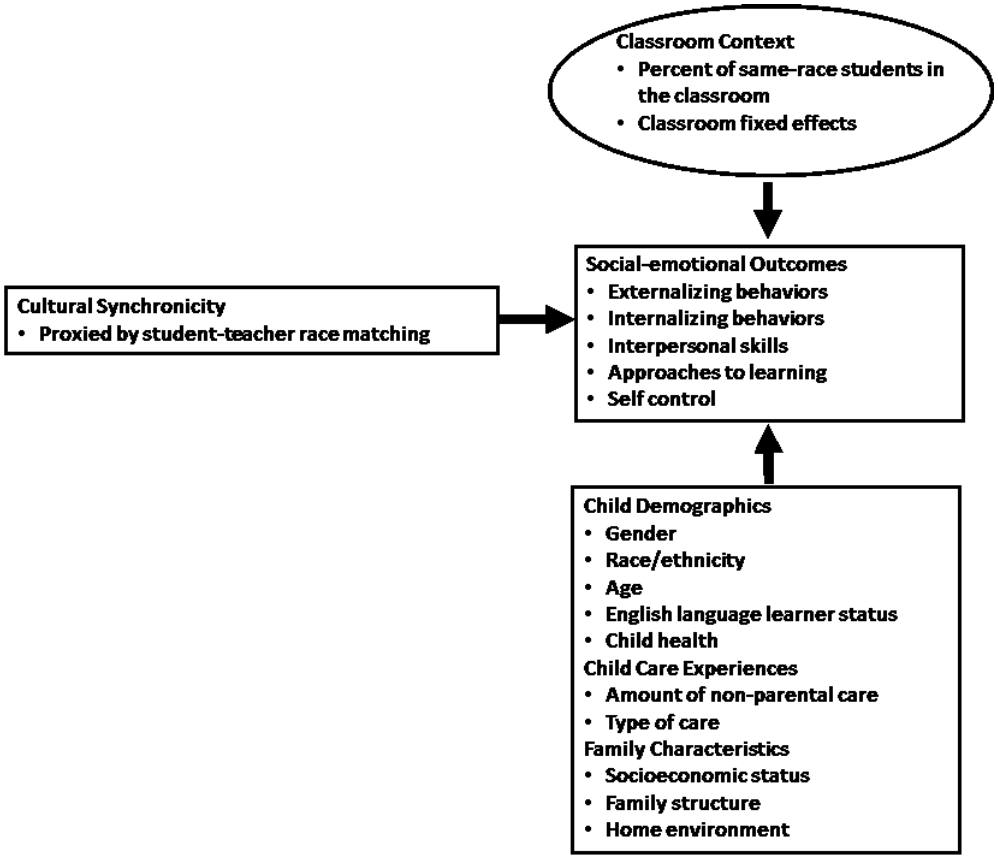

Figure 1 – Theoretical Model Linking Cultural Synchronicity to Social-Emotional Outcomes

Chapter 3

Estimating Spillovers of Teachers from Selective Certification Programs: The Case of Teach For America

3.1 Introduction

Recent concerns regarding teacher shortages and quality have prompted states to reevaluate their teacher certification requirements. A long-established barrier to teaching in public schools in the United States, teacher licenses are state-specific and usually require completion of a teacher preparation program, passing certification exams, and a certain amount of student-teaching experience. Alternative teacher certification (AC) programs aim to lower the barriers to earning teacher certification for individuals who have already obtained at least a bachelor's degree, allowing these programs to tap new pools of potential teachers. Many AC programs place teachers in schools with a history of underperformance. It is in these schools that the presence of a new type of teacher may be particularly helpful in changing the status quo. This study focuses on the impact of one of the most selective AC programs, Teach For America (TFA).[39]

Since it was founded in 1989 to address issues of educational inequality in low-income communities, TFA has become one of the most prominent AC programs in the United States. TFA aims to identify and place high quality teachers, primarily recent college graduates, who are willing to teach for two years in low-income communities across the nation. While a number of studies have assessed the impact of TFA teachers on their own students, the larger effect of TFA (and other selective AC programs) on placement schools is an important open question.

The goal of this study is to determine whether TFA teachers affect student performance in other grades. Using panel data of grade-level test scores from California elementary schools along with TFA placement data, I analyze the effect of TFA placement

[39] Full disclosure: I was a teacher with Teach For America from 2008-2010.

on math and English scores in grades adjacent to TFA grades (e.g., looking at third and fifth grade when a TFA teacher is present in fourth grade). Ideally, I would use student or classroom-level data to determine whether TFA affects student performance within the same grade. Lacking this level of data, I investigate whether TFA affects adjacent grades. My primary empirical strategy is a difference-in-differences estimation that exploits variation in the timing and location of TFA teacher placement. My findings suggest that grades adjacent to TFA grades improve as TFA presence increases. The size of the impact when TFA constitutes a relatively large percentage of teachers is roughly 10% of a standard deviation in both math and English. This translates into moving students up in the grade-year score distribution by about 4 percentile points in both subjects on average. I find no statistically significant change in TFA-adjacent grade scores when TFA has a relatively small presence.

There are a few studies that examine the influence of TFA teachers, but these studies primarily focus on a TFA teacher's own students. These studies compare the students of TFA teachers with the students of their non-TFA same-grade colleagues (i.e., comparing teachers in the same grade, school, and year). In the only experimental study of TFA teachers, where both teachers and students were randomly assigned to classrooms for one year, Glazerman, Mayer, and Decker (2006) find that TFA teachers have no impact on reading achievement and a positive impact on math achievement. The relative gains of TFA students in math are 15% of a standard deviation and roughly equivalent to one extra month of instruction. When compared to other novice teachers (three or fewer years of experience), the math achievement gain rises to 26% of a standard deviation. Non-experimental evaluations of TFA teachers, whose results rely on being able to control for the nonrandom sorting of teachers and students to classrooms, find that students of TFA teachers do

approximately the same as those of regular new teachers (e.g., Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2006; Kane, Rockoff, & Staiger, 2008; Boyd, Grossman, Hammerness, Lankford, Loeb, Ronfeldt et al., 2012).

Critics of alternative certification programs often contend that teachers from these programs ultimately harm student achievement through higher turnover rates (Darling-Hammond, Holtzman, Gatlin, & Heilig, 2007). Due to their explicit two-year commitment, turnover is indeed high among TFA teachers. Kane et al. (2008) measure the five-year attrition rate among TFA teachers to be 82%, compared to 50% for traditionally certified teachers. They find that despite a large difference in retention rate between TFA and traditionally certified teachers, the steady-state impacts of TFA (allowing for different initial effectiveness and returns to experience) were no different from traditionally certified teachers.

A key identification assumption of comparing teachers within a grade, as the aforementioned studies of TFA do, is that there are no within-grade teacher spillovers. That is, a teacher has no effect on his colleagues' students. Jackson and Bruegmann (2009), however, find that students have larger achievement gains in math and reading when their teacher has more effective same-grade colleagues. In order to avoid the reflection problem (Manski, 1993), teacher quality is determined by out-of-sample value-added estimates and observable characteristics that are not affected by contemporaneous peer quality (namely, experience and teacher certification exam scores). Their findings suggest that for a teacher with three colleagues, replacing one of his colleagues with a one standard deviation better teacher has about 20% of the effect on student test scores as replacing the own teacher with a one standard deviation better teacher. Finding that spillovers are strongest for

inexperienced teachers and tend to persist, the authors attribute the measured spillovers to peer learning.

Though it is plausible that teacher spillovers extend across grades, Jackson and Bruegmann (2009) only establish spillovers within the same grade.[40] There is, however, evidence of peer quality affecting worker productivity in various contexts, even when worker output is independent. For example, Mas and Moretti (2009) show that checkout clerks work faster when viewing high productivity workers.[41] In addition, peer salience can have real effects as well. Bandiera, Barankay, and Rasul (2010) find that the productivity of socially-tied workers converges only when they work alongside one another, even though compensation is independent. These findings suggest that peers can affect one another in more indirect ways by altering social incentives.

How might TFA teachers affect teachers in adjacent grades? Jackson and Bruegmann (2009) offer three sources of spillovers between teachers: joint production and shared resources, motivation and effort, and peer learning. In an across-grade setting, the most plausible of these mechanisms is motivation and effort. TFA teachers may motivate their colleagues through shared enthusiasm, performance comparison, or out of fear of losing their jobs to the following year's class of incoming TFA recruits. The entrance of TFA into a school may signal that the school has new source of teachers, essentially decreasing the employment protection for existing teachers.[42] As Jacob (2013) demonstrates, decreasing

[40] The authors include a school-year fixed effect (instead of the typical school-grade-year fixed effect) so that they are comparing teachers in the same school and year. They state that their narrow definition of peers – teachers within the same grade – therefore provides a lower bound of the estimate of peer importance.

[41] See also Ichino and Maggi (2000).

[42] Some superintendents have been forthright with their intention to keep TFA teachers amidst layoffs to regular teachers. In a 2009 USA Today article, Peter Forman,

employment protection in public schools can lead to increased teacher productivity. Conversely, if TFA teachers are less skilled or motivated than teachers in adjacent grades, the entrance of TFA may also have a negative effect.

Identifying the causal effects of TFA presence on adjacent grades presents an empirical challenge as TFA teachers are not randomly assigned to schools. Schools that partner with TFA may have recently suffered a setback in student test scores and taken comprehensive corrective action or may have an enthusiastic principal and would have experienced improved student outcomes in the absence of TFA. My difference-in-differences empirical strategy accounts for measured time-varying school characteristics that may affect both the hiring of TFA teachers and student achievement. I also repeat my preferred specifications, which compare only schools within the same school district, with a matched sample to help ameliorate concerns of non-random TFA placement across schools. The main findings are robust to a number of alternative specifications. Additionally, there is no evidence of student sorting between TFA and non-TFA schools or that the results are driven by contemporaneous school-level factors.

This paper sheds light on the broader impact of teachers from selective AC programs such as TFA. Specifically, I find that the presence of this type of teacher has implications for teachers in other grades within a school. Furthermore, in establishing spillovers across grades, this work has significant implications for value-added models and constructing appropriate counterfactual groups. Although my data do not support analysis of the

superintendent of Charlotte-Mecklenburg, North Carolina schools was reported to have told the school board that despite massive layoffs of non-TFA teachers, 100 TFA teachers would be retained due to prior commitments (Toppo, 2009).

mechanisms that generate the observed spillovers, this work provides a first step towards understanding the across-grade influence of teachers within a school.

The organization of the remainder of this paper is as follows. Section 3.2 provides a brief overview of Teach For America. Section 3.3 discusses the data sources used in the analysis. Section 3.4 details the empirical strategy. Section 3.5 presents results, a sensitivity analysis, and investigates alternative explanations for the primary findings. Section 3.6 concludes.

3.2 Teach For America

Originally based on the Princeton University undergraduate thesis of founder Wendy Kopp, TFA is a non-profit organization that aims to eliminate educational inequality by recruiting highly skilled men and women to teach for two years in low-income schools. Admission into TFA is very competitive; in 2011, 5,200 "corps members" were selected from approximately 48,000 applications (Johnson, 2011). The majority of corps members lack formal training in education but are much more likely than non-TFA teachers in TFA placement schools to have attended top undergraduate schools.[43]

After being accepted into TFA and before entering placement schools, corps members take part in a five-week intensive training that focuses on teaching and curriculum development. TFA provides ongoing support to its classroom teachers in the form of professional development opportunities, mentoring, and instructional resources. TFA works

[43] Glazerman et al. (2006) report that the percentage of teachers having attended "competitive" colleges in their sample was 70 percent for TFA teachers and less than 3 percent for non-TFA teachers. However, there is little evidence that teacher academic background is a good predictor of student success (e.g., Clotfelter et al., 2006; 2007; Harris & Sass, 2011).

to balance district needs with corps member preferences in determining where teachers are ultimately placed.

Although small in terms of the percentage of all teachers in the United States, TFA teachers comprised roughly six percent of all alternatively certified teachers in the 2008-2009 school year (National Center for Alternative Certification, 2010). In the 2011-2012 school year, more than 9,000 TFA teachers in 43 regions taught over 600,000 students, typically in Title I schools in rural and urban areas.[44] Corps members are hired and paid directly by the school district in which they teach and compensation is typically the same as that of regular first year teachers. Most districts pay TFA about $1,500 for each new teacher to offset the cost of training and recruiting (Dobbie, 2011).

3.3 Data

Data for the empirical analysis come from merging California Department of Education school and grade-level data with TFA placement records. To measure student achievement, I use publicly available grade-level average scores on the California Standards Tests as well as the Academic Performance Index, which is a school-level measure of performance on all of the California Standards Tests and is used as the state's accountability measure.[45] Data provided by TFA are from Los Angeles and the Bay Area in California and include the number of teachers by subject, grade, school, and year.

[44] In order to meet Title I eligibility, at least 40 percent of a school's students must be from low-income families, where low-income is defined by the United States Census.

[45] The California Standards Tests, a part of the California Standardized Testing and Reporting (STAR) Program, are taken each Spring by students in grades two through eleven. Test data and additional information are available online through the California Department of Education's website, http://www.cde.ca.gov. The STAR program is expected to be replaced by the Smarter Balanced Assessment System in the 2014-2015 school year.

I focus on math and English scores of third-grade through fifth-grade students from 2003-2011.[46] TFA grades have at least one TFA teacher from 2007-2011 who teaches math, English, or both. Since I only observe the placement of TFA teachers while they are in TFA (i.e., during their first two years), I exclude from my sample grades that received a TFA teacher any time before 2007 and after I no longer observe a TFA teacher in a grade.[47] Overall, there are 109 TFA-grade observations in English and 104 for math for grades three through five.

Potential spillover grades are based on TFA placement that meets the above criteria.[48] I label grades that are adjacent to TFA grades as spillover grades in the same year. For example, if TFA is placed in third grade in a school in year *t*, I designate grade four as a spillover grade in year *t*. In order to eliminate the possibility of mistaking a lagged TFA own-grade effect for a spillover effect, the sample of spillover grades does not include student-grade cohorts that have previously been exposed to a TFA teacher.[49] In total, there are 57 adjacent-grade observations in English and 55 in math. Appendix Table 3.1 lists the school districts and number of unique schools in the sample with at least one TFA teacher or

[46] These are the grades Jackson and Bruegmann (2009) use to establish teacher peer effects. California Standards Tests data are available as early as second grade, though using second grade in the analysis would not allow me to control for previous grade outcomes. 2002 was the first year the California Standards Tests were administered.

[47] This is to avoid mislabeling a grade that contains a former TFA teacher. I have TFA placement data from 2001-2011.

[48] I consider TFA placement in second through sixth grade to construct the sample of third, fourth, and fifth adjacent-to-TFA grades.

[49] For example, say TFA teaches third grade in years *t* and *t+1* and fourth grade is labeled as a spillover grade in those same years. Any estimated spillover effect in year *t+1* might just be a delayed direct effect of TFA. Although this significantly reduces my sample of spillover grades, results are similar if these grades are included.

one spillover grade (I label these as simply "TFA schools") in grades three through five. These data come from 50 different schools in nine school districts.

Because corps members are typically placed in schools with a high proportion of students from low-income families, I base my control group of schools on the percent of students in a school that are eligible for free and reduced-price lunch in the first year that TFA is observed in a district.[50] Specifically, I restrict my control sample to include elementary schools in TFA school districts that have no prior TFA placement in any grade and a percent of students eligible for free and reduced-price lunch that is at least the minimum of TFA schools within the district. The minimum threshold is based on the first year that TFA appears in a district (instead of all years) to prevent a few outlier TFA school observations that appear later in the sample to drive the control group. For example, in 2009, a TFA teacher is placed in a school in the Los Angeles Unified School District (LAUSD) with 0% free and reduced-price lunch (the average of TFA schools in LAUSD is 73%). I reexamine my main specification with free and reduced-price lunch outliers dropped in a robustness check below.

Table 3.1 gives school-level characteristics by TFA school status any time during the sample period. School-level covariates used in my analysis are: percent of students who are black, percent of students who are Hispanic, percent of students who are white, percent of students eligible for free and reduced-price lunch, percent of students who are English language learners, average parent education level, and student enrollment. These time-varying characteristics are used to capture factors that affect both student performance and

[50] Conversations with TFA's director of district and school partnerships indicate that in California TFA targets schools that have at least 60% of students eligible for free and reduced-price lunch.

are associated with the hiring of TFA teachers.[51] TFA schools tend to have a higher percentage of black and English language learner students, a lower percentage of Hispanic and white students, a lower percentage of students eligible for free and reduced-price lunch, and smaller school enrollment. Because of these differences, I examine the robustness of my results to using a matched sample based on school-level variables.

The primary dependent variable in the analysis is the grade-level mean score from the California Standards Tests in either math or English. As these tests are not vertically scaled (i.e., designed to be compared across grades), each score is converted to a Z-score (mean zero and unit variance) based on scores of all tested California students by subject, grade, and year. This process is repeated for the Academic Performance Index, where mean and variance measurements are based on school-year totals. This is done so that the effects measure how scores move around in the overall distribution of student achievement. Table 3.1 shows that the Academic Performance Index at TFA schools is lower than non-TFA schools on average.

Table 3.2 provides grade-level student performance and enrollment by subject and TFA grade-level status any time during the sample period. Compared to grades in non-TFA schools, TFA grades have on average lower math and English scores. TFA-adjacent grades also have a lower mean math score than non-TFA schools, but their English scores are nearly identical. The descriptive statistics for student grade enrollment, also used as a covariate, confirm that TFA schools tend to be smaller than non-TFA schools.

[51] Except for grade-level student enrollment, demographic data are not available at the grade level. Thus, school-level characteristics are used as a proxy for grade-level characteristics.

3.4 Empirical Strategy

In its most basic form, the identification strategy is to use a difference-in-differences estimator. In order to assess the impact of TFA on adjacent-grade test scores, control grades from non-TFA schools are used to represent what would have happened in the absence of TFA. Looking before and after TFA is present, a comparison of the change in test scores in the adjacent grade relative to the change in an unweighted average of the test scores in the control grades yields an estimate of the effect of TFA across grades.

Since I have more than one time period in which school grades receive TFA, I employ a more generalized fixed-effect specification. Grade-level student performance (the mean score in math or English), Y, is modeled as

(3.1) $$Y_{gst} = \beta_0 + \beta_1 TFA_{gst} + \beta_2 AdjTFA_{gst} + \boldsymbol{X'_{gst}\beta_3} + \beta_4 Y_{(g-1)s(t-1)} + \phi_{dgt} + \lambda_{gs} + \varepsilon_{gst}$$

where g is grade, s is school, t is year, and d is district. The variable *TFA* is an indicator for the presence of TFA. $AdjTFA$ indicates whether the grade observation is adjacent to a TFA grade.[52] The coefficient of interest, β_2, gives the effect of being adjacent to a TFA grade on grade-level scores. The vector X is comprised of the time-varying school and grade characteristics described above.[53] The lagged achievement of the student cohort is included to control for student quality.[54] The district-grade-year fixed effect, ϕ, controls for time effects that are common across district grades; it ensures that comparisons are made between schools in the same district, grade, and year. That is, it assumes the appropriate control

[52] Though I am only interested in the effect of TFA presence on adjacent grades, I include the own-grade effect for completeness.

[53] This specification contains grade enrollment (in logs), not school enrollment.

[54] For example, if the outcome variable was 4th grade math scores in 2010, this variable would be the 3rd grade math scores in 2009.

grades for a TFA-adjacent grade are all the same grades within the same district (an assumption that will be relaxed later). The school-grade fixed effect, λ, controls for school-specific grade effects that are time-invariant, or the fixed quality of a school grade. Standard errors are clustered at the school level as same-school observations are unlikely to be independent (Bertrand, Duflo, & Mullainathan, 2004).

The empirical evidence on spillovers suggests that in addition to peer quality, peer salience can have substantial effects. Thus, I allow the effect of TFA teachers on adjacent-grade outcomes to vary by amount of presence. I use teacher count data from the California Department of Education to measure the number of full-time equivalent math and English teachers in TFA-adjacent grades.[55] I define "TFA presence" for any given TFA-adjacent grade as the number of TFA teachers who are one grade removed from the TFA-adjacent grade divided by the number of teachers in the TFA-adjacent grade.[56] TFA presence is considered high if it is greater than or equal to 50% and low if it less than 50% (but greater than zero). Due to a large number of TFA presence observations at 50%, this yields 34 (of the possible 55) adjacent-grade observations with high TFA presence math and 35 (of the possible 57) in English. I redistribute observations from high to low to equalize these categories in a sensitivity analysis below.

In my preferred specification, β_2 from equation (3.1) is divided into the aforementioned high and low-TFA-presence categories. The identifying assumption for

[55] These data are not available for the 2009-10 school year. As the number of teachers rarely changes from year to year within a grade, I interpolate the 2009-10 school year from the preceding year and subsequent year.

[56] Although access to teacher count data yields an estimate of TFA presence within a grade, within-grade spillovers are not identifiable. The reason for this is explained in Appendix A.3.

these coefficients is that being labeled a TFA-adjacent grade is random conditional on the time-varying school and grade characteristics, lagged achievement of the student cohort, and fixed effects. Testing this assumption involves comparing trends in test scores in TFA-adjacent grades and control grades in the years preceding intervention, which is done in the results section below.

Including school-grade fixed effects alleviates concerns that grades adjacent to TFA grades are fundamentally better or worse than control grades. School-grade fixed effects cannot, however, control for unobserved time-varying factors that affect both student performance and TFA placement. Also, schools may hire TFA teachers in response to declining test scores or test scores may have been trending upward before TFA placement. I include school-grade time trends to address these issues.

Another potential source of bias stems from school-level changes that occur simultaneously with TFA placement. One possibility that may confound the estimate of the TFA spillover effect is that a principal brings in TFA teachers as part of a larger plan to increase student test scores. If contemporaneous with the presence of TFA, these school-level changes would make it impossible to separately identify the TFA spillover effect. To investigate this issue, I examine school-level changes in student performance when TFA is present. School-level student performance, measured by its Academic Performance Index (API), is modeled as

(3.2) $$API_{st} = \alpha_0 + \alpha_1 TFA_{st} + \boldsymbol{X'_{st}\alpha_2} + \omega_{dt} + \nu_s + \varepsilon_{st}.$$

The TFA presence measure, TFA, is binary in the basic specification but is allowed to vary based on TFA presence in the school. The vector X contains the same time-varying school characteristics as before except school enrollment is included instead of grade enrollment. A

district-year fixed effect, ω, is included so that comparisons are made between schools within the same district and year. The school fixed effect, ν, controls for any time-invariant school effects that might affect TFA placement and API. Standard errors are clustered at the school level.

3.5 Results

3.5.1 Main Results

I first test the common trend assumption that is imperative for the difference-in-differences framework in equation (3.1). Figure 3.1 presents results from two regressions, the top panels give the math results and the bottom panels give the English results, which show how scores in TFA-adjacent grades change over time relative to control grades. Grade-level scores are regressed on a set of indicator variables for years before TFA placement by TFA presence (high or low) along with school-grade fixed effects and district-grade-year fixed effects.[57] The omitted category is three years or more before TFA placement. The 90% confidence interval is formed from standard errors clustered at the school level.

None of the lead indicators are statistically different from zero in Figure 3.1. This suggests TFA-adjacent grades and the control grades in non-TFA schools share similar pre-treatment trends in test scores. In high-TFA-presence grades, there is an increase in math and English scores in the year TFA is present and the coefficients are statistically different

[57] There are not a sufficient number of observations to separately examine the second year that TFA is placed in a school grade. Also, there are no "lagged" observations (years after TFA is present) as adjacent TFA grades are dropped from the sample after the TFA teacher is no longer observed as being present.

from zero. There appears to be no effect of low TFA presence on adjacent grades in either subject.

Table 3.3 presents baseline results from equation (3.1).[58] Each column represents the results from a separate regression. Columns (1) and (3) include a binary measure for grades adjacent to TFA in math and English, respectively. There appears to be no effect of TFA placement on adjacent grades in math and a positive and significant effect in English. Columns (2) and (4) contain results of the preferred specification where the TFA-adjacent-grade indicator is separated into high and low TFA presence. The results suggest that a relatively large presence of TFA in an adjacent grade lead to a roughly 10% of a standard deviation increase in both math and English. These results represent about a four percentile point increase from the mean scores in control schools for both subjects (math: -0.149 + 0.1, English: -0.265 + 0.1). There is no effect on adjacent grades when TFA constitutes a relatively small percentage of teachers.

3.5.2 Sensitivity Analysis

The first robustness check I perform provides insight into the nature of student achievement in the years prior to TFA placement. The goal is to see whether exam scores change before the arrival of TFA. If this is the case, then perhaps academic interventions before TFA is first placed are responsible for any observed treatment effect. Table 3.4 presents a model that includes two lead indicators and the year of TFA presence with the omitted category being three years or more before TFA is placed. Separate indicators are

[58] I discuss only the point estimates on adjacent TFA grades. Results with TFA own-grade estimates are presented in Appendix Table 3.2.

created for high versus low-TFA-presence grades. This model extends the one used to produce Figure 3.1 by including school and grade covariates. Including leads does not significantly alter the estimates from Table 3.3. No individual lead is statistically significant in either model and a joint test of significance of both leads for each TFA presence category cannot reject they are jointly equal to zero.

I examine a number of additional specifications to test the sensitivity of the math results in Table 3.5. Column (1) reports the baseline results from the preferred specification found in the second column of Table 3.3. Column (2) presents results from a model with school-grade linear trends added to equation (1). These trends capture unobserved factors that change smoothly over time at the school-grade level. The high-TFA-presence estimate in column (1) remains positive but is less precisely estimated and is not significantly different from zero at conventional levels. The coefficient on adjacent grades that are exposed to a relatively low TFA presence remains statistically indistinguishable from zero.

Table 3.1 demonstrates that TFA and non-TFA schools differ across many observed characteristics. Thus, I re-estimate the preferred specification with a propensity matched subset of schools.[59] To match TFA schools with observably similar control schools, a propensity score for each observation is obtained from a probit model that regresses TFA placement at the school level on the same time-varying school covariates used in equation (3.2). Appendix Table 3.3 gives the results of the probit regression. TFA teachers tend to be placed in schools with a higher percentage of black, Hispanic, and English language learner students, a smaller percentage of students eligible for free and reduced-price lunch, and

[59] See Becker and Ichino (2002) for more information on using propensity score matching.

smaller enrollment. Each observation is stratified into a block based its propensity score. I then only consider looking between schools that are in the same block. This amounts to replacing the district-grade-year fixed effect in equation (3.1) with a block-grade-year fixed effect. 109 control school observations are omitted because they are not matched with any TFA schools.

Column (3) of Table 3.5 presents the results from estimating equation (3.1) with a propensity matched subsample. On average, high-TFA-presence grades experience increases in math scores and these results are significant at the 1% level. The coefficient magnitudes are similar to those in the baseline. The coefficient on low-TFA-presence grades is once again small and not statistically significant.

In column (4), a regression of the preferred specification is weighted by grade enrollment. Weighted least squares allows for the estimated impact on larger grades to be given more weight. The high-TFA-presence coefficient is statistically significant at the 10% level and equal 10% of a standard deviation.

To make sure that a few TFA-adjacent-grade observations from schools with very low levels of free and reduced-price lunch are not driving the results, I drop the lowest 10% of these observations in column (5) of Table 3.5. The high-TFA-presence estimate remains large but, as expected, is measured with less precision.

A potential worry is that the imbalance in the number of observations for high and low TFA presence is affecting the results. Thus, I relabel the six high-TFA-presence grades with the largest grade enrollments as "low TFA presence" in both math and English.[60]

[60] This yields 28 (of the possible 55) observations of high-TFA-presence grades in math and 29 (of the possible 57) in English.

Estimates of equation (3.1) under this labeling system are given in column (6). The results change little from the baseline specification.

The corresponding sensitivity analysis for the English results is given in Table 3.6. In general, the English results are less sensitive to modeling assumptions than the math results. The high-TFA-presence coefficients in columns (2) through (6) are statistically significant and similar in magnitude to the baseline estimate. The estimated effect on low-TFA-presence grades is not statistically different from zero in any model.

3.5.3 Alternative Explanations

My results suggest that grades adjacent to TFA grades improve when TFA teachers constitute a relatively large percentage of teachers. Because all of the specifications involve comparing grades across schools, it remains that some unobserved school-level factor that is correlated with TFA presence and student performance could be driving the results. This would imply that there is some other school-level change positively affecting student scores occurring at the same time as TFA placement, but only at schools with high-TFA-presence grades. If this is true, positive effects should be evident at the school level in these schools.

To test this, I scale all observations to the school-year level and create an indicator variable for whether the school contains at least one grade in a given year that has been labeled as a high-TFA-presence grade in either Math or English. There are 30 school-year observations in the sample that meet this criterion. Table 3.7 provides results from equation (2.2), which regresses a school's Academic Performance Index (API) on TFA indicators, school variables, a school fixed effect, and a district-year fixed effect. The regression in column (1) includes a single indicator for whether TFA is present in a given school year.

The estimated coefficient is positive but not statistically significant. In column (2), the TFA indicator is broken down into separate indicators: whether the school contains at least one high-TFA-presence grade, and all other TFA schools. While the coefficient on this first indicator is positive (as to be expected since a school's API is a function of grade-level scores) it is not significantly different from zero.

Lastly, I test for another type of potential bias: student sorting. Any positive effect of TFA placement could be discredited if, upon learning of TFA placement, parents move their children into TFA schools. If children from these families perform systematically better than other students, the impact of TFA spillovers will be biased upward. To test for student sorting, I look for changes in average parent education and the percent of students eligible for free and reduced-price lunch in schools that contain at least one high-TFA-presence grade.[61] Table 3.8 presents results from regressing average parent education, in column (1), and percent of students eligible for free and reduced-price lunch, column (2), on dummies for whether the school contains at least one high-TFA-presence grade and all other TFA schools. These regressions also include school fixed effects and district-year fixed effects. Importantly, the dummy for the school having at least one high-TFA-presence grade is not statistically different from zero in either model.

3.6 Conclusion

[61] Average parent education and the percent of students eligible for free and reduced-price lunch (FRPL) are highly correlated with student performance in my sample. Using math scores, the correlation is .318 and -.262 for parent education and FRPL, respectively. Using English scores, the correlation is .531 and -.437 for parent education and FRPL, respectively.

Despite the growing relevance of AC programs, little is known about the broader impact teachers from these programs have. This study focuses on Teach For America and exploits variation in the timing and location of TFA teacher placement to identify the effect of TFA teachers on adjacent-grade outcomes. Using data from the California Department of Education and TFA, I find that grades next to TFA grades improve as TFA teachers become more salient. Specifically, grades that experience at least a 50% TFA presence, defined as the number of TFA teachers in adjacent grades divided by the number of teachers in the own grade, improve by roughly four percentile points in the grade-year score distribution in both math and English. The estimated effect on math scores (10% of a standard deviation) is roughly 40-66% of the size of the previously estimated effect of a TFA teacher on his own class scores.

While previous studies have examined the effect of teachers on their same-grade colleagues, this is the first study to investigate the across-grade influence of a type of teacher that might be particularly effective at changing the social incentives within a school. My findings imply that empiricists need to be judicious with their model assumptions when estimating the impact of teachers; including a school-year fixed effect may not be appropriate if spillovers are occurring across grades.

Due to data limitations, this work should be seen as a modest first step towards understanding how teachers from selective AC programs affect their colleagues. Since I measure student performance at the grade level and am not able to control for individual teacher quality, a possible confounding factor would be if sorting of non-TFA teachers is correlated with TFA placement. For example, if the hiring of high-quality regular teachers and TFA placement occur simultaneously, then it might appear that TFA placement affects

student achievement in adjacent grades. However, in order to undo my findings, this could only be the case when TFA teachers represent a relatively high percentage of teachers as I find no spillovers when TFA constitute a low percentage of teachers.

Though this work contributes to understanding spillovers across grades, it cannot distinguish among the possible mechanisms of how teachers in different grades might affect one another. Future work should determine whether mechanisms such as performance comparison or employment protection are contributing factors to teacher productivity when TFA teachers are present. Having to utilize school grades that receive TFA for the first time, I am likely capturing the initial effect of TFA placement on adjacent grades. Whether spillover effects persist as TFA colleagues become more accustomed to TFA presence is left to future research. Also, if TFA teachers are inducing improvements across grades, then it is likely that they are also affecting teachers within their own grade. Establishing and understanding TFA spillovers within a grade would provide a much clearer picture as to the greater impact of this program and other selective AC programs like it.

References (Chapter 3)

Bandiera, O., Barankay, I., & Rasul, I. (2010). Social incentives in the workplace. *The Review of Economic Studies*, *77*(2), 417-458.

Becker, S. O., & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. *Stata Journal*, *2*(4), 358-377.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in- differences estimates? *The Quarterly Journal of Economics*, *119*(1), 249-275.

Boyd, D., Grossman, P., Hammerness, K., Lankford, H., Loeb, S., Ronfeldt, M., & Wyckoff, J. (2012). Recruiting effective math teachers: Evidence from New York City. *American Educational Research Journal*, *49*(6), 1008-1047.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education Finance and Policy*, *1*(2), 176-216.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources*, *41*(4), 778-820.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, *26*(6), 673-682.

Darling-Hammond, L., Holtzman, D., Gatlin, S., & Heilig, J. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education Policy Analysis Archives*, *13*(42), 1-48.

Dobbie, W. (2011). Teacher characteristics and student achievement: Evidence from Teach For America. Unpublished manuscript.

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

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

Ichino, A., & Maggi, G. (2000). Work environment and individual background: Explaining regional shirking differentials in a large Italian firm. *The Quarterly Journal of Economics*, *115*(3), 1057-1090.

Jackson, C. K., & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics*, *1*(4), 85-108.

Jacob, B. A. (2013). "The effect of employment protection on teacher effort." *Journal of Labor Economics*, *31*, 727-761.

Johnson, J. Teach for America 2011 acceptance rate: 11 percent. The Washington Post 3 August 2011. <http://www.washingtonpost.com/blogs/campus-overload/post/teach-for-america-2011-acceptance-rate-11-percent/2011/08/03/gIQAqX8bsI_blog.html>, Accessed on February 27, 2014.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, *27*(6), 615-631.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, *60*(3), 531-542.

Mas, A., & Moretti, E. (2009). Peers at Work. *American Economic Review*, *99*(1): 112-145.

National Center for Alternative Certification. Introduction and overview. <http://www.teach-now.org/intro.cfm>, Accessed on April 1, 2013.

Toppo, G. Teach For America: Elite corps or costing older teachers jobs? USA Today 29 July 2009. <http://usatoday30.usatoday.com/news/education/2009-07-29-teach-for-america_N.htm>, Accessed on June 1, 2013.

Tables (Chapter 3)

Table 3.1 – School-Level Characteristics by TFA School Status

	TFA school		Non-TFA school		Difference in means
	Mean	SD	Mean	SD	
Academic Performance Index Z-score	-0.68	0.05	-0.51	0.02	-0.17***
Percent black students	18.62	22.12	16.16	22.91	2.47*
Percent Hispanic students	62.80	26.10	66.62	29.05	-3.82**
Percent white students	6.42	12.84	4.40	7.96	2.02***
Percent free and reduced-price lunch	76.35	22.49	84.04	15.37	-7.68***
Percent English language learners	47.53	20.79	44.78	19.74	2.76**
Average parent education level	2.17	0.57	2.19	0.45	-0.01
Student enrollment	333.02	113.19	416.98	235.68	-83.97***
N	315		2538		

Difference in means statistical significance is given at the ***1%, **5%, and *10% levels.

Notes: All means are unweighted averages. The unit of observation is school-year. TFA schools receive a TFA teacher at least once from 2007-2011. Non-TFA schools never receive a TFA teacher during the sample period. Parent education is coded as follows: 1 – not high school graduate; 2 – high school graduate; 3 – some college; 4 – college graduate; 5 – some graduate school and higher.

Table 3.2 – Grade-Level Student Performance and Enrollment by TFA Grade Status

	Adjacent-to-TFA grade		TFA school grade		Non-TFA school grade	
	Mean	SD	Mean	SD	Mean	SD
Math Z-score	-0.231	0.024	-0.305	0.019	-0.149	0.005
Score diff. from non-TFA	-0.082***		-0.156***			
Grade enrollment	65.46	3.12	72.83	2.47	98.16	0.63
N math	280		433		6733	
English Z-score	-0.264	0.012	-0.332	0.016	-0.265	0.004
Score diff. from non-TFA	0.001		-0.067***			
Grade enrollment	65.85	3.05	73.72	2.42	98.16	0.62
N English	291		444		6733	

Difference in means statistical significance is given at the ***1%, **5%, and *10% levels.

Notes: All means are unweighted averages. The unit of observation is grade-school-year. The "adjacent-to-TFA grade" designation is for grades that are at any time one grade away from the nearest TFA grade. "TFA school grade" are grades that receive a TFA teacher at any time during the sample period. Non-TFA school grades are grades in schools that never receive a TFA teacher during the sample period. "Score diff. from non-TFA" is the difference between the category Z-score mean and the non-TFA school Z-score mean.

Table 3.3 – TFA Presence and Grade-Level Student Performance: Baseline Results

	Math		English	
	(1)	(2)	(3)	(4)
Adjacent-to-TFA grade	0.064		0.081**	
	(0.046)		(0.034)	
Adjacent-to-TFA grade: high TFA presence		0.104*		0.110**
		(0.056)		(0.044)
Adjacent-to-TFA grade: low TFA presence		-0.001		0.037
		(0.073)		(0.046)
R^2	0.776	0.776	0.841	0.841
N	7541	7541	7573	7573
School-grade fixed effects	Yes	Yes	Yes	Yes
District-grade-year fixed effects	Yes	Yes	Yes	Yes
School and grade variables	Yes	Yes	Yes	Yes
Lagged cohort score	Yes	Yes	Yes	Yes

Statistical significance is given at the ***1%, **5%, and *10% levels.

Notes: Each column reports coefficients from a separate OLS regression. The dependent variable is the math or English Z-score for a given grade, school, and year. Standard errors are clustered at school level and are in parentheses. The "adjacent-to-TFA grade" designation is for grades that are one grade away from the nearest TFA grade. TFA presence is defined as the number of TFA teachers who are one grade removed from the TFA-adjacent grade divided by the number of teachers in the TFA-adjacent grade and is considered high if it is at least 50% and low if it is less than 50%. Columns (1) and (3) include a single indicator for adjacency to TFA grades. Columns (2) and (4) break down the adjacent-to-TFA-grade indicator into high TFA presence and low TFA presence. School and grade variables are percent of students that are black, percent of students that are white, percent of students that are Hispanic, percent of students that are eligible for free and reduced-price lunch, percent of students that are English language learners, log of grade enrollment, and average parent education. The lagged cohort score is the math or English score of the current grade-level students in the previous grade.

Table 3.4 – TFA Presence and Grade-Level Student Performance: Timing of TFA Introduction

	Math (1)	English (2)
Two years before TFA Adjacent-to-TFA grade: **high** TFA presence	0.042 (0.062)	-0.001 (0.053)
One year before TFA Adjacent-to-TFA grade: **high** TFA presence	0.058 (0.052)	-0.002 (0.051)
Year of TFA Adjacent-to-TFA grade: **high** TFA presence	0.120* (0.070)	0.111** (0.055)
Two years before TFA Adjacent-to-TFA grade: **low** TFA presence	-0.044 (0.082)	0.038 (0.057)
One year before TFA Adjacent-to-TFA grade: **low** TFA presence	-0.031 (0.082)	0.076 (0.057)
Year of TFA Adjacent-to-TFA grade: **low** TFA presence	-0.013 (0.062)	0.058 (0.057)
R^2	0.776	0.841
N	7541	7573
School-grade fixed effects	Yes	Yes
District-grade-year fixed effects	Yes	Yes
School and grade variables	Yes	Yes
Lagged cohort score	Yes	Yes

Statistical significance is given at the ***1%, **5%, and *10% levels.

Notes: Each column reports coefficients from a separate OLS regression. The dependent variable is the math or English Z-score for a given grade, school, and year. The omitted category is 3+ years before TFA first entered the school. Standard errors are clustered at school level and are in parentheses. The "adjacent-to-TFA grade" designation is for grades that are one grade away from the nearest TFA grade. TFA presence is defined as the number of TFA teachers who are one grade removed from the TFA-adjacent grade divided by the number of teachers in the TFA-adjacent grade and is considered high if it is at least 50% and low if it is less than 50%. School and grade variables are percent of students that are black, percent of students that are white, percent of students that are Hispanic, percent of students that are eligible for free and reduced-price lunch, percent of students that are English language learners, log of grade enrollment, and average parent education. The lagged cohort score is the math or English score of the current grade-level students in the previous grade.

Table 3.5 – TFA Presence and Grade-Level Student Performance: Sensitivity Analysis of Math Results

	Baseline	Time trends	Matched sample	Weighted regression	Drop FRPL outliers	Balance treated obs.
	(1)	(2)	(3)	(4)	(5)	(6)
Adjacent-to-TFA grade: high TFA presence	0.104*	0.082	0.114**	0.100*	0.101	0.110*
	(0.056)	(0.066)	(0.052)	(0.055)	(0.063)	(0.062)
Adjacent-to-TFA grade: low TFA presence	-0.001	0.008	-0.034	-0.011	-0.012	0.018
	(0.073)	(0.023)	(0.075)	(0.071)	(0.076)	(0.063)
R^2	0.776	0.846	0.770	0.788	0.776	0.776
N	7541	7541	7432	7541	7536	7541
School-grade fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
District-grade-year fixed effects	Yes	Yes	No	Yes	Yes	Yes
Block-grade-year fixed effects	No	No	Yes	No	No	No
School and grade variables	Yes	Yes	Yes	Yes	Yes	Yes
Lagged cohort score	Yes	Yes	Yes	Yes	Yes	Yes

Statistical significance is given at the ***1%, **5%, and *10% levels.

Notes: Each column reports coefficients from a separate OLS regression. The dependent variable is the math Z-score for a given grade, school, and year. Standard errors are clustered at school level and are in parentheses. The "adjacent-to-TFA grade" designation is for grades that are one grade away from the nearest TFA grade. TFA presence is defined as the number of TFA teachers who are one grade removed from the TFA-adjacent grade divided by the number of teachers in the TFA-adjacent grade and is considered high if it is at least 50% and low if it is less than 50%. Column (1) gives the baseline estimates from the second column of Table 3.3. Column (2) includes school-grade time trend. Column (3) re-estimates the baseline specification on matched samples designated by "block." The matched sample regressions have fewer observations because 109 control school observations are not matched with any TFA schools. In column (4), the regression is weighted by grade enrollment. Column (5) drops the bottom 10% of adjacent TFA grades in terms of free and reduced-price lunch. I relabel the six high-TFA-presence grades with the largest grade enrollments as "low TFA presence" in column (6). School and grade variables are percent of students that are black, percent of students that are white, percent of students that are Hispanic, percent of students that are eligible for free and reduced-price lunch, percent of students that are English language learners, log of grade enrollment, and average parent education. The lagged cohort score is the math score of the current grade-level students in the previous grade.

Table 3.6 – TFA Presence and Grade-Level Student Performance: Sensitivity Analysis of English Results

	Baseline	Time trends	Matched sample	Weighted regression	Drop FRPL outliers	Balance treated obs.
	(1)	(2)	(3)	(4)	(5)	(6)
Adjacent-to-TFA grade: high TFA presence	0.110**	0.125***	0.124***	0.101**	0.090*	0.111**
	(0.044)	(0.042)	(0.042)	(0.041)	(0.046)	(0.051)
Adjacent-to-TFA grade: low TFA presence	0.037	0.002	0.006	0.031	0.014	0.052
	(0.046)	(0.054)	(0.046)	(0.052)	(0.043)	(0.041)
R^2	0.841	0.884	0.837	0.860	0.841	0.841
N	7573	7573	7464	7573	7568	7573
School-grade fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
District-grade-year fixed effects	Yes	Yes	No	Yes	Yes	Yes
Block-grade-year fixed effects	No	No	Yes	No	No	No
School and grade variables	Yes	Yes	Yes	Yes	Yes	Yes
Lagged cohort score	Yes	Yes	Yes	Yes	Yes	Yes

Statistical significance is given at the ***1%, **5%, and *10% levels.

Notes: Each column reports coefficients from a separate OLS regression. The dependent variable is the English Z-score for a given grade, school, and year. Standard errors are clustered at school level and are in parentheses. The "adjacent-to-TFA grade" designation is for grades that are one grade away from the nearest TFA grade. TFA presence is defined as the number of TFA teachers who are one grade removed from the TFA-adjacent grade divided by the number of teachers in the TFA-adjacent grade and is considered high if it is at least 50% and low if it is less than 50%. Column (1) gives the baseline estimates from the fourth column of Table 3.3. Column (2) includes school-grade time trend. Column (3) re-estimates the baseline specification on matched samples designated by "block." The matched sample regressions have fewer observations because 109 control school observations are not matched with any TFA schools. In column (4), the regression is weighted by grade enrollment. Column (5) drops the bottom 10% of adjacent TFA grades in terms of free and reduced-price lunch. I relabel the six high-TFA-presence grades with the largest grade enrollments as "low TFA presence" in column (6). School and grade variables are percent of students that are black, percent of students that are white, percent of students that are Hispanic, percent of students that are eligible for free and reduced-price lunch, percent of students that are English language learners, log of grade enrollment, and average parent education. The lagged cohort score is the English score of the current grade-level students in the previous grade.

Table 3.7 – TFA Presence and School-Level Student Performance

	Academic Performance Index	
	(1)	(2)
TFA	0.084	
	(0.075)	
Contains at least one high-TFA-presence grade		0.112
		(0.086)
R^2		0.881
N	2843	2843
School fixed effects	Yes	Yes
District-year fixed effects	Yes	Yes
School variables	Yes	Yes

Statistical significance is given at the ***1%, **5%, and *10% levels.

Notes: Each column reports coefficients from a separate OLS regression. The dependent variable is the Academic Performance Index Z-score for a given school and year. Standard errors are clustered at school level and are in parentheses. Column (1) includes a single indicator for TFA being present in a given school in a given year. In column (2) The "contains at least one high-TFA-presence" designation is a school-level indicator variable that is 1 for schools that have at least one grade in a given year that is a high-TFA-presence grade in either math or English and 0 otherwise. School variables are percent of students that are black, percent of students that are white, percent of students that are Hispanic, percent of students that are eligible for free and reduced-price lunch, percent of students that are English language learners, log of student enrollment, and average parent education.

Table 3.8 – Testing for Student Sorting

	Parent education	% free and reduced-price lunch
	(1)	(2)
Contains at least one high-TFA-presence grade	-0.061	-0.241
	(0.040)	(3.03)
R^2	0.907	0.829
N	2843	2843
School fixed effects	Yes	Yes
District-year fixed effects	Yes	Yes
School variables	No	No

Statistical significance is given at the ***1%, **5%, and *10% levels.

Notes: Each column reports coefficients from a separate OLS regression. The dependent variable is average parent education in column (1) and the percent of students eligible for free and reduced-price lunch in column (2). Both dependent variables are measured at the school-year level. Standard errors are clustered at school level and are in parentheses. The "contains at least one high-TFA-presence grade" designation is a school-level indicator variable that is 1 for schools that have at least one grade in a given year that is a high-TFA-presence grade in either math or English and 0 otherwise.

Figures (Chapter 3)

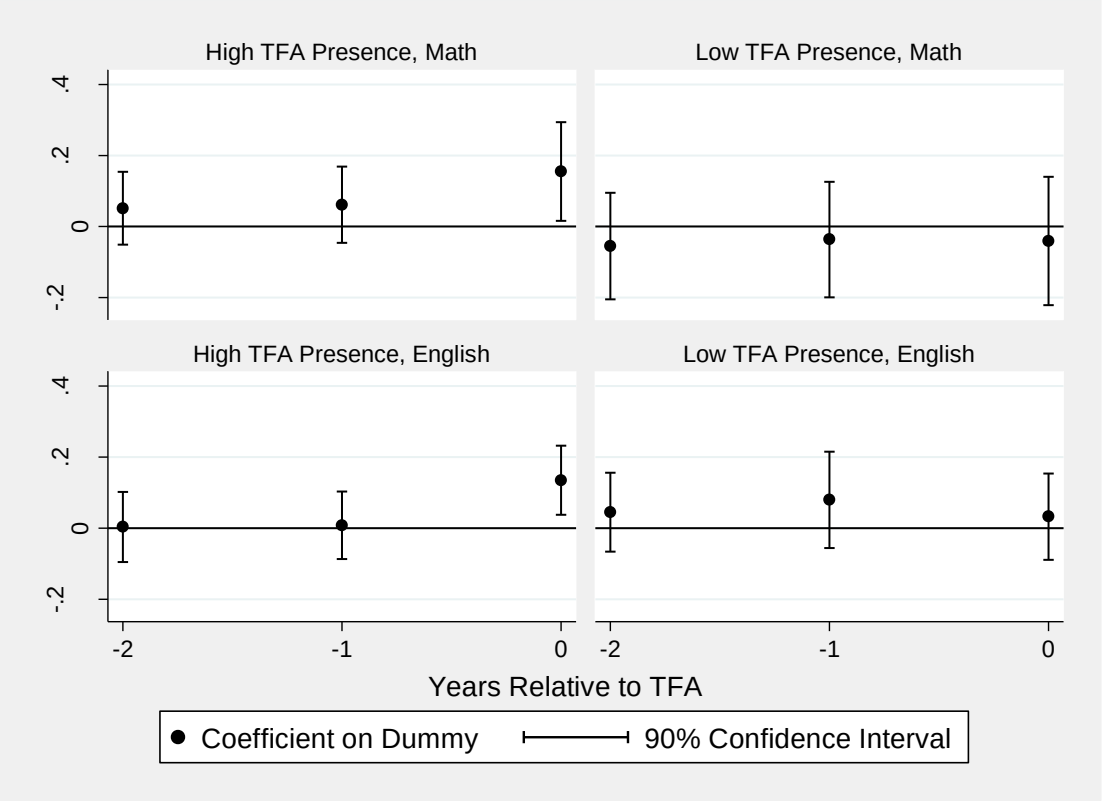

Figure 3.1 – Math and English Grade-Level Performance by TFA Presence

Notes: This is a plot of TFA-adjacent-grade coefficients from two regressions, one for math and one for English, of grade-level test scores on a set of indicator variables for years before TFA placement by TFA presence (high or low) along with school-grade fixed effects and district-grade-year fixed effects. The omitted category is three or more years before TFA first entered the school. The top panels give the math results and the bottom panels give the English results. The 90% confidence interval is formed from standard errors clustered at the school level. TFA presence is defined by the number of teachers in the TFA grade divided by number of teachers in the adjacent grade and is considered high if it is at least 50% and low if it is less than 50%.

Appendices

Appendix to Chapter 1

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:1999 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:1999 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:1999 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

Appendix to Chapter 2

A.2 Two Threats to Validity

One threat to identifying the relationship between student-teacher race match for persons of color and student outcomes stemmed from nonrandom student sorting to classrooms. Our estimates of δ_1 and δ_2 in equation (2.2) may have been biased if unobserved student quality was correlated with both the teacher-student interaction for persons of color and our outcome variables. For example, highly motivated students of color may have sorted into classrooms with a teacher of color. Since we compared students within a classroom where there was a teacher of color to students within a classroom where there not a teacher of color, we examined this threat to validity by testing whether gaps in observable student characteristics that are plausibly correlated with unobserved student ability were different in classrooms where there was a teacher of color compared to classrooms where there was not a teacher of color. We modeled student characteristic x_{ic} as

(A.2.1) $$x_{ic} = \lambda_0 + \lambda_1 PCstud_i + \lambda_2 PCteach_c * PCstud_i * same_{ic} + \lambda_3 PCteach_c * PCstud_i * diff_{ic} + \omega_c + \nu_{ic}.$$

We estimated λ_2 and λ_3, which describe the difference in minority gaps in x_{ic} across minority and nonminority teacher classrooms, for the relevant student covariates in equation (2.2).

Next, we addressed another form of omitted-variable bias: teacher bias. In contrast to student achievement, which can be assessed with standardized measures, assessments of social-emotional skills were obtained through teacher reports, which may be subjected to bias. Thus, any observed gains in teacher-scored social-emotional skills for students of color matched with teachers of color may simply be due to teacher race-based observer bias.

Previous work has shown that conditional on (objective) test scores, teachers assessed same-race students more favorably in math and reading (Ouazad, 2014). In order to test for teacher bias, we took advantage of the fact that the ECLS-K:2011 contains both test scores and subjective teacher assessments of student ability in math and reading. Unfortunately, the data do not contain test-based social-emotional scores to compare with the teacher assessments of social-emotional skills. Thus, we followed Ouazad (2014) and modeled the math or reading teacher survey-based assessment fall-spring gain score of student *i* in classroom *c* as

(A.2.2) $$TA_{ic} = \sigma_0 + \sigma_1 PCstud_i + \sigma_2 PCteach_c * PCstud_i * same_{ic}$$
$$+\sigma_3 PCteach_c * PCstud_i * diff_{ic} + \sigma_4 TS_{ic} + \psi_c + \boldsymbol{X_i'\theta} + u_{ic}$$

which controls for the appropriate standardized test score fall-spring gain in math or reading, TS_{ic}. Conditional on the (objective) test score, one would expect there to be no effect of race or ethnicity interactions in the absence of teacher race-based bias. That is, we expected σ_2 and σ_3 to be statistically different from zero only in the presence of teacher bias. We also modeled the dependent variable as the difference between the teacher survey-based assessment and the test score for each subject, similar to Lavy (2008). Despite these precautions, an absence of teacher bias in math or reading assessments may not translate to social-emotional outcomes; assessing math and reading ability may be easier and less prone to race-based bias than behavioral measures. For example, the behaviors of students of color may be interpreted as acting out by white, whereas teachers of color, by virtue of their shared cultural background, may not perceive the same behaviors as necessarily disruptive (Alexander, Entwisle, & Thomson, 1987). Since our gain-score model used the difference in spring and fall teacher assessments, any constant teacher bias will be differenced out. In

other words, as long as teacher bias in assessing social-emotional outcomes was consistent over time, our results were not biased due to teacher perceptions.

We tested for differential student sorting by estimating equation (A.2.1) with student characteristics that are plausibly correlated with unobserved student ability. Appendix Table 2.1 gives both coefficients for the person of color teacher-student interaction: the same-race coefficient in the first column and the different-race coefficient in the second column. Only the same-race and different-race coefficients for center-based care were statistically different from zero at conventional levels, suggesting that differences in observable characteristics between students of color and white students in general did not depend on teacher race. If anything, having relatively less center-based care in the year before kindergarten for students of color matched with teachers of color would attenuate any observed positive effect of race match on outcomes.

Next, we tested for teacher bias in assessing math and reading scores. Appendix Table 2.2 presents the results of estimating equation (A.2.2), providing coefficient estimates for the two person of color teacher-student interactions. Estimates from the specification with the outcome as the difference between the teacher assessment and test score for each subject are given in columns (1) and (2). Regressions represented in columns (3) and (4) had the subject-specific teacher assessment as the dependent variable and controlled for the corresponding test score. Columns (2) and (4) added student controls to their respective baseline models. In no model did we detect evidence of teacher bias for the growth scores in math or reading.

Appendix Table 2.1 – Testing for Student Sorting across Classroom Types

	Person of color teacher-student interaction	
	Same race	Different race
Outcome: Student demographic characteristics		
Male	0.04	0.01
	(0.07)	(0.09)
Age at kindergarten entry (months)	0.39	0.24
	(0.59)	(0.61)
English Language Learner	-0.05	0.03
	(0.04)	(0.06)
Physical health scale	0.02	0.10
	(0.12)	(0.12)
Outcome: Student care		
Hours of care in year before kindergarten	0.94	-2.18
	(2.11)	(2.41)
Relative care in year before kindergarten	0.01	-0.02
	(0.06)	(0.07)
Non-relative care in year before kindergarten	0.00	-0.02
	(0.05)	(0.06)
Center-based care in year before kindergarten	-0.11*	-0.15**
	(0.06)	(0.07)
Never attended center-based care	-0.00	0.10
	(0.06)	(0.07)
Head Start in year before kindergarten	-0.00	-0.06
	(0.05)	(0.05)
Outcome: Family characteristics		
Socioeconomic status composite	0.00	-0.13
	(0.10)	(0.10)
Two biological parents at home	-0.00	0.06
	(0.07)	(0.07)
Parental involvement composite	0.19	-0.04
	(0.15)	(0.18)
Home learning activities composite	0.10	0.15
	(0.15)	(0.15)

Notes. Each row represents a separate OLS regression. Standard errors were clustered at the class level and observations were weighted using ECLS-K:2011 sample weights (W12T0).
*p < 0.1, **p < 0.05, ***p<0.01

Appendix Table 2.2 – Testing for Teacher Bias in Math and Reading

	(1)	(2)	(3)	(4)
Subject: Math				
Observations: 8,970				
Person of color teacher-student interaction:	0.02	-0.01	-0.00	-0.01
same race	(0.14)	(0.14)	(0.12)	(0.12)
Person of color teacher-student interaction:	0.02	-0.02	-0.05	0.04
different race	(0.14)	(0.14)	(0.13)	(0.13)
Subject: Reading				
Observations: 8,780				
Person of color teacher-student interaction:	0.14	0.12	0.08	0.07
same race	(0.12)	(0.12)	(0.10)	(0.10)
Person of color teacher-student interaction:	-0.05	-0.05	-0.06	-0.06
different race	(0.12)	(0.12)	(0.12)	(0.11)
Fixed effects				
Classroom	Yes	Yes	Yes	Yes
Controls				
Student controls	No	Yes	No	Yes
Subject test score	No	No	Yes	Yes

Notes. Each pair of same race and different race teacher-student interactions represents a separate OLS regression. Standard errors were clustered at the class level and observations were weighted using ECLS-K:2011 sample weights (W12T0).
*p < 0.1, **p < 0.05, ***p<0.01

Appendix to Chapter 3

A.3 The Detection of Within-Grade TFA Spillovers

A natural place to start when determining the extent of TFA spillovers would be to examine whether TFA teachers affect their same-grade colleagues. Unfortunately, observing grade-level outcomes makes identification of within-grade spillovers infeasible. To see this, assume the grade level score (Y) is a weighted average of TFA and non-TFA classroom scores (TFA and $nonTFA$, respectively), which can be written as

(A.3.1) $$Y = TFA * \frac{n_{TFA}}{N_{grade}} + nonTFA * \frac{n_{nonTFA}}{N_{grade}}$$

where N_{grade} is the number of teachers in the grade, n_{TFA} is the number of TFA teachers, and n_{nonTFA} is the number of non-TFA teachers. Assume that the non-TFA classroom score is a function of the percentage of TFA teachers in a grade such that $nonTFA = S\left(\frac{n_{TFA}}{N_{grade}}\right)$. The goal is to find $dS/d\left(\frac{n_{TFA}}{N_{grade}}\right)$. The observed variation in equation (A1) is the grade-level test score and percentage of the grade that is TFA teachers. What is not identifiable is how TFA compares to $nonTFA$. Thus, if one observes that Y increases as the percentage of TFA teachers increases, it cannot be ruled out that this is simply due to $TFA > nonTFA$. Similarly, if one observes that Y decreases as the percentage of TFA teachers increases, it cannot be ruled out that this is simply due to $TFA < nonTFA$. In either of these cases, the presence of TFA spillovers within a grade cannot be cleanly identified.

Appendix Table 3.1 – School Districts Included in the Sample

School district	Region	Number of TFA schools
Alum Rock Union Elementary	Bay Area	4
Compton Unified	Los Angeles	1
Franklin-McKinley	Bay Area	1
Los Angeles Unified	Los Angeles	10
Oakland Unified	Bay Area	10
Ravenswood City	Bay Area	4
San Francisco Unified	Bay Area	2
San Jose Unified	Bay Area	11
West Contra Costa Unified	Bay Area	7

Notes: The number of TFA schools is the number of unique schools in the sample that have at least one TFA teacher or at least one spillover grade from 2007-2011.

Appendix Table 3.2 –TFA Presence and Grade-Level Student Performance: Baseline Results with TFA Grade

	Math		English	
	(1)	(2)	(3)	(4)
Adjacent-to-TFA grade	0.064		0.081**	
	(0.046)		(0.034)	
Adjacent-to-TFA grade: high TFA presence		0.104*		0.110**
		(0.056)		(0.044)
Adjacent-to-TFA grade: low TFA presence		-0.001		0.037
		(0.073)		(0.046)
TFA grade	0.084*	0.083*	0.063**	0.063**
	(0.044)	(0.044)	(0.026)	(0.026)
R^2	0.776	0.776	0.841	0.841
N	7541	7541	7573	7573
School-grade fixed effects	Yes	Yes	Yes	Yes
District-grade-year fixed effects	Yes	Yes	Yes	Yes
School and grade variables	Yes	Yes	Yes	Yes
Lagged cohort score	Yes	Yes	Yes	Yes

Statistical significance is given at the ***1%, **5%, and *10% levels.

Notes: Each column reports coefficients from a separate OLS regression. The dependent variable is the math or English Z-score for a given grade, school, and year. Standard errors are clustered at school level and are in parentheses. The "adjacent-to-TFA grade" designation is for grades that are one grade away from the nearest TFA grade. "TFA grade" are grades that receive a TFA teacher. TFA presence is defined as the number of TFA teachers who are one grade removed from the TFA-adjacent grade divided by the number of teachers in the TFA-adjacent grade and is considered high if it is at least 50% and low if it is less than 50%. Columns (1) and (3) include a single indicator for adjacency to TFA grades. Columns (2) and (4) break down the adjacent-to-TFA-grade indicator into high TFA presence and low TFA presence. School and grade variables are percent of students that are black, percent of students that are white, percent of students that are Hispanic, percent of students that are eligible for free and reduced-price lunch, percent of students that are English language learners, log of grade enrollment, and average parent education. The lagged cohort score is the math or English score of the current grade-level students in the previous grade.

Appendix Table 3.3 – Probit Model for Propensity Score Analysis

	TFA school
Percent black students	0.010**
	(0.004)
Percent Hispanic students	0.010***
	(0.004)
Percent white students	0.008
	(0.008)
Percent free and reduced-price lunch	-0.016***
	(0.003)
Percent English language learners	0.009**
	(0.004)
Average parent education	-0.207
	(0.181)
Log(enrollment)	-0.405***
	(0.106)
Pseudo R^2	0.067
N	2843

Statistical significance is given at the ***1%, **5%, and *10% levels.

Notes: This are results from a probit regression where the dependent variable is the probability of a school receiving at least one TFA teacher in a given year. Standard errors are in parentheses.