Disproportionate Discipline and Racial Differences in Educational Attainment: The Effect of Suspension on College Matriculation

by

Claire E. Kunesh

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Education

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Susan D. Holloway, Chair

Professor Sophia Rabe-Hesketh

Professor Jasjeet S. Sekhon

Fall 2017

Abstract

Disproportionate Discipline and Racial Differences in Educational Attainment: The Effect of Suspension on College Matriculation

by

Claire E. Kunesh

Doctor of Philosophy in Education

University of California, Berkeley

Professor Susan D. Holloway, Chair

Professor Sophia Rabe-Hesketh

Professor Jasjeet S. Sekhon

In this study, I use multilevel modeling and multivariate matching to estimate the effect of suspension on college attendance. I build upon existing research that has suggested a negative relationship between school suspension and educational outcomes but that has mostly not been able to control for pre-existing differences in academic achievement between suspended and non-suspended students. I also extend my analysis to college attendance, which is currently an under-studied outcome related to high school suspension and thought to be important for many life outcomes. After controlling for baseline academic achievement, socioeconomic status, and other potential confounders, the college attendance rate of students who were suspended once or twice in the first semester of tenth grade was approximately 10 percentage points lower than that of non-suspended students. These results suggest that racial differences in educational attainment are in part due to the disproportionate discipline of African American students.

This dissertation is dedicated to my grandfathers Charles and Warren, who first taught me about race, education, and power in America.

Contents

List of Tables

List of Figures

Acknowledgements

I benefited from the help of many individuals during my doctoral career at Berkeley. Foremost, I'd like to acknowledge the guidance and support of my dissertation committee, Jas, Sophia, and Susan. I also thank Amity Noltemeyer, who encouraged me to pursure my PhD and first formally introduced me to the issues at the heart of this dissertation; my father, who set me off on my academic journey by teaching me to read and pushing me to excel academically; my mother, who instilled a critical perspective in me from a young age; Alex Hancock, who supported me in countless ways and whom I'll never be able to properly thank in words; Andreas Strand, who was instrumental in presenting an initial version of part of the matching proportion of this project in Jas Sekhon's seminar; Yotam Shem-Tov, Andrew McCall, and other members of Jas Sekhon's research group who provided me with helpful feedback; Amy Arneson, James Mason, and other members of Sophia Rabe-Hesketh's research group; the members of Susan Holloway's research group, especially Becky Hachmyer and Charissa Tansomboon, who provided me with invaluable advice at important decision points; and Jessica Ernandes Naecker, Anna Casey, Christine Gerchow, and Chloe Green, who graciously welcomed me into their cohort. My experience at Berkeley and my dissertation itself was shaped by my experience in Cristina Mora's class on race and ethnicity; I'm so grateful for her and my classmates' seminar meetings. I'd also like to thank the many students at College Track who shared their stories with me, especially Betel Yimer, Lauryn Lu, and Kobe Sellers; they also shaped this dissertation and my doctoral studies. Finally, I'd like to thank the hundreds of Cal undergraduates whom I had the pleasure of teaching. Their warm presence in class balanced the hours of solitary work on my research, and their energy and work ethic motivated me.

Chapter 1

Introduction

The United States educational system is characterized by stark disparities in educational attainment. While the highest-performing high school students graduate and go on to attend some of the most highly ranked universities in the world, the lowest-performing high school students drop out or graduate only having achieved basic math and literacy skills that are not sufficient to succeed professionally in the modern economy. Frequently, these disparities are racialized; White students are much more likely to be part of the former group whereas Black[1] students are more likely to be part of the latter (Vanneman, Hamilton, Anderson, & Rahman, 2009). The academic achievement gap between White and Black American students has been highly studied but is not yet fully understood. In this dissertation, I examine how exclusionary discipline is associated with educational attainment to assess whether the disproportionate suspension of African American students contributes to this gap in educational outcomes.

Exclusionary discipline is commonly used in American schools, and African American students are generally two to three times more likely to be referred to the office and suspended than White students (D. J. Losen, 2011; Skiba et al., 2011; Wallace, Goodkind, Wallace, & Bachman, 2008). Exclusionary discipline is a potentially harmful form of punishment because of its very nature – students are removed from their classrooms and miss valuable instruction time. Therefore, they have a reduced opportunity to learn while they are excluded from class and potentially thereafter if they have missed foundational skills or knowledge upon which subsequent instruction relies. Excluded students may be labelled as troublemakers by

[1]The terms Black and African American are used interchangeably in this paper. These labels are used in accordance with federal data collection and the vast majority of published research, although these categories obfuscate substantial within-group variation. Some students who identify as African American or Black achieve at high levels, and no one factor can explain the lower achievement of those who do not. Within all ethnic/racial groups, there is significant variation in family backgrounds, educational experiences, and life outcomes. Moreover, there is substantial overlap between the educational outcomes of different ethnic/racial groups. In general, within-group variation exceeds between-group variation. For example, in a nationally-representative study of children in kindergarten and first grade, race, socioeconomic status, and gender combined explained only 1–8% of the variation in learning rates (Downey, von Hippel, & Broh, 2004). However, I focus on the Black-White gap in educational attainment because this average difference is large and is significant in terms of life outcomes.

teachers; this labelling may affect teachers' attributions for future behaviors and their academic expectations for such students. Such labels may also lead students to associate with peers who have academic difficulties or engage in maladaptive behaviors. Being excluded for disciplinary reasons may also affect students' academic motivation and their relationships with teachers if students perceive that they have been treated unfairly.

If exclusionary discipline is causally related to poorer academic outcomes, this would constitute an important social justice issue and pose an economic concern to the United States as a whole. This link would constitute an ethical, social justice issue because Black students are more likely to be suspended than White students for reasons other than differences in behavior. This link would constitute an economic concern because there is an unmet need for highly educated, skilled workers in the United States. Additionally, people with lower levels of education are more likely to become involved in the criminal justice system and have poorer health outcomes, both of which are costly to the American government and taxpayers (Levin, 2009).

This project builds upon past research that has found that students who are suspended tend to earn lower grades and test scores and are more likely to be retained in a grade and drop out than those who are not suspended (see review by Noltemeyer, Ward, & Mcloughlin, 2015). Much of the existing correlational research suggests a relationship but is potentially subject to confounding because it has not controlled for baseline academic achievement or other potential confounders. Clearly, being removed from school may result in academic difficulties. However, students who have pre-existing academic problems may be more likely to be suspended. Other research suggests a relationship but is limited by its reliance on school-level variables (e.g., school-level dropout rates and average academic achievement at single time point). This school-level research is potentially subject to the ecological fallacy – relationships at the aggregate level may not reflect relationships at the individual level – in addition to reverse causality or confounding. Given the economic research that argues suspension may benefit non-suspended students' achievement and that the racial achievement gap would improve if schools serving African American students' enacted harsher discipline policies (Kinsler, 2013), it is vital to assess the consequences of suspension using more rigorous causal inference methods. This analysis contributes to the existing literature by matching students based on their initial academic performance and comparing college matriculation outcomes four years later, which is currently an understudied outcome with respect to the consequences of suspension.

1.1 Research Questions

1. After accounting for potential confounds such as socioeconomic status and baseline differences in academic achievement, are students who received one or two in-school or out-of-school suspensions less likely to attend a four-year college or university than students who were not suspended?
2. If so, is this reduced likelihood related to any of the following factors: increased likelihood of dropping out from high school, lower subsequent math achievement (after

controlling for baseline differences), associating with peers who do not plan to attend college, decreased likelihood of taking college entrance exams, being less likely to apply to college, or being accepted by fewer colleges?

I accounted for variables that are thought to be related both to suspension and college attendance (or some prerequisite to college attendance) based on empirical research or theory and that were indeed related to both suspension and college attendance in this data set. As described in greater detail below, these variables include history of retention (e.g. Fabelo et al., 2011; Jimerson, Anderson, & Whipple, 2002), living in a single parent household (e.g. Bowers, Sprott, & Taff, 2012; Morris & Perry, 2016), having an identified disability (e.g. Newman, Wagner, Cameto, Knokey, & Shaver, 2010; Sullivan, Klingbeil, & Van Norman, 2013), race/ethnicity (e.g. Perna & Titus, 2005; Skiba et al., 2011), academic achievement (e.g. Bowditch, 1993; Kupchik, 2010), gender (e.g. Goldin, Katz, & Kuziemko, 2006; Wallace et al., 2008), socioeconomic status (e.g. Caro, Cortina, & Eccles, 2015; Feliciano & Ashtiani, 2012; Sullivan et al., 2013), and school characteristics (e.g. Hill, 2008; Skiba et al., 2014).

Given that existing correlational research and psychoeducational theory suggest a negative relationship between suspension and educational outcomes, I hypothesized that suspended students would be less likely to attend college than students who were not suspended after conditioning on these variables. However, because the majority of existing research in education has not controlled for baseline differences in academic achievement and has not examined college matriculation as on outcome, I did not have strong prior beliefs about the size of this association. My analysis of potential mediating factors was exploratory and based on prior correlational research in education (e.g., Noltemeyer et al., 2015), an existing survey of college admissions workers (Weissman & NaPier, 2015), and theory about the stigmatizing effects of exclusionary discipline that is grounded in qualitative research (e.g., Bowditch, 1993). Given the lack of prior research comparing the influence of such potential mediators, I made no hypotheses about their relative strength. This conceptual model is illustrated in Figure 6.1.

I used two analytical perspectives to examine my primary research question: multilevel modeling and matching. Following the recommendations of Morgan and Winship (2015), I used multiple matching methods. I matched individuals based on a univariate propensity score and through multivariate genetic and coarsened exact matching. I conducted my analysis using a longitudinal data set of 10,799 students attending public high schools across the United States.

Chapter 2

The Path to College Attendance

In 2015, approximately 93% of non-Hispanic White Americans and 89% of Asian Americans aged 25 years or older had at least a high school diploma or General Educational Development (GED) certificate of equivalency, which is necessary to pursue a post-secondary education, compared to 87% of Black Americans (Ryan & Bauman, 2016). It should be noted that Black Americans are much more likely than White Americans to earn a GED rather than a high school diploma (GED Testing Service, 2014; Heckman & LaFontaine, 2010); although nominally equivalent, a GED may be viewed as less attractive than a high school diploma to potential employers and is associated with much worse labor market outcomes (Heckman, Humphries, & Mader, 2010). This high-school-education gap seems to be stable or improving among young Americans. Among 25- to 29-year-olds in 2014, 96% of White Americans and 97% of Asian/Pacific Islander Americans completed at least a high school diploma or its equivalent compared to 92% of Black Americans (Kena et al., 2015). In other words, across all American adults 25 years or older, the White-Black gap in high school graduation rates was 6%, but among 25- to 29-year olds the gap was 4%. On the other hand, the gap in attaining a college education appears to be worsening. Approximately 33% of non-Hispanic White Americans and 54% of Asian Americans aged 25 years or older had at least a bachelor's degree compared to 23% of Black Americans (Ryan & Bauman, 2016). Among 25- to 29-year-olds, 41% of White Americans and 61% of Asian/Pacific Islander Americans had earned at least a bachelor's degree compared to only 22% of Black Americans (Kena et al., 2015). In fact, the White-Black gap among 25- to 29-year-olds in attaining a bachelor's degree widened from 13 to 18 percentage points between 1990 and 2014.

2.1 Reasons for Racial Differences in Academic Achievement and Educational Attainment in the United States

These gaps have sometimes been attributed to differences in student motivation or attitudes. For example, some have proposed that African American students' low academic achievement

is due to their rejection of academic success as "acting white," in response to discrimination and perceived limited opportunities (e.g., Ogbu, 1987). Although some teachers do attribute differences in achievement to students' motivation, beliefs, and "culture" (e.g., Bol & Berry, 2005), this oppositional culture hypothesis is not supported by much empirical research. For example, in a quantitative study of adolescents living in Maryland ($N = 1,480$), African American participants reported higher perceived returns to education and higher educational aspirations than White participants before and after controlling for background characteristics (Harris, 2006). In a national study of students enrolled in Grade 11 in 2012, 63% of Black students aspired to graduate from college or graduate school compared to 65% of White students (Schneider & Saw, 2016).

Qualitative research also suggests that African American adolescents, even those who live in poverty, still embrace education as a path to success (Carter, 2005). In her qualitative study of 68 low-income adolescents living in Yonkers, New York in the 1990s, Carter (2005) found that these young people defined success as "finish[ing] school and college" (p. 20). For example, 100% of the African American participants and 92% of the Latino participants agreed that "achievement and effort in school lead to job success later on" (p. 113). Nevertheless, some students, whom Carter termed "noncompliant believers," could not or chose not to adapt to dominant behavioral norms such as style of speech and dress. Some of the participants in Carter's study resisted behaviors conducive to school achievement (while still endorsing the value of education), but this was due to gender or the intersection of race and gender rather than race itself. For example, speaking Standard English was derided as feminine by some of the male participants (p. 84). The adherence to White, middle-class norms (such as style of dress) may indirectly affect these adolescents' achievement by impacting teachers' perceptions of these students. In another qualitative study of middle-income high school students, no African American student mentioned "acting white" or described academic success using this term (Lewis & Diamond, 2015). Overall, these findings contradict widespread stereotypes that position African American adolescents as oppositional.

Other explanations for the racial gap in academic achievement and educational attainment include parent involvement, socioeconomic status, school resources, and teacher quality, which are frequently interrelated constructs. While a comprehensive review of these bodies of literature is beyond the scope of this dissertation, I will briefly address some of their major hypotheses and findings. Although African American parents, on average, report that they have high aspirations for their children and are involved in their lives (Robinson & Harris, 2014), their involvement may sometimes be less effective than that of White parents. When African American parents do attempt to advocate for their children in school settings as White parents do, these attempts are frequently viewed negatively by school staff and are consequently less successful (Lareau & Horvat, 1999; Lewis & Diamond, 2015). Second, as I discuss below, African American children are more likely to be raised by parents with lower levels of education and lower incomes. This may affect the quality of the academic support and material resources parents can provide to their children.

Race and socioeconomic status are closely intertwined in American society. For example, people of low socioeconomic status are more likely to self-identify and be perceived by others as African American (Freeman, Penner, Saperstein, Scheutz, & Ambady, 2011; Saperstein &

Penner, 2012).[1]. African Americans also face obstacles to discrimination in the job market (e.g., Pager, 2003) that lead to lower socioeconomic status. Consequently, African American children grow up in households that have, on average, significantly lower incomes and wealth than White Americans. Therefore, although African American and White parents appear to be similarly involved in their children's education at home, many African American parents are less able than many White parents to support their children's education in financial and material ways. These economic differences are especially important given the amount of time American children spend outside of school. Although American children do not spend a low amount of time in school compared to children in other major economies (DeSilver, 2014), the United States is characterized by greater amounts of socioeconomic inequality than most post-industrial democracies.

As top incomes have risen and lower ones have remained stagnant, the gap between high- and low-income households' expenditures on children has grown substantially (Duncan & Murnane, 2011). In the early 1970s, households in the top income quintile spent approximately four times as much on their children's "enrichment" than bottom income quintile homes. Since that time, low-income households expenditures have risen but have not kept pace with the increase in high-income households' expenditures. In the early 2000s, the top income households spent almost seven times as much as low-income households. These and other socioeconomic differences in children's early environments lead to differences in educational preparedness between rich and poor children and White and African American children. Due to their households' lower socioeconomic status (Fryer & Levitt, 2004), on average, African American kindergartners enter school less prepared than their White counterparts, and they perform worse on tests of math and reading than White kindergartners. These early deficits may interfere with future learning, exacerbating initial differences (Duncan & Magnuson, 2011). Early math and reading achievement predicts achievement in later grades, even after accounting for the differences in socioeconomic status and cognitive ability that are associated with initial gaps (Duncan et al., 2007).

A substantial amount of the variation in educational outcomes appears to be related to unequal opportunities outside of school, as gaps exist at the beginning of kindergarten. Additionally, the gaps between high- and low-performing children tend to grow during the summer months (Downey et al., 2004). However, research on the "summer setback" suggests that the school environment in fact aggravates racial gaps in achievement. For example, in a nationally representative study of children in kindergarten and first grade, Downey et al. (2004) found that on a test of reading achievement "a standard deviation's advantage in [socioeconomic status] predict[ed] a relative gain of .16 points per month during summer, but only .07 points per month during kindergarten and .05 points per month during first grade" (where the average learning rate was 1.65 points per month during kindergarten; p. 624). In other words, the socioeconomic status gap was greater during summer months than during the school year. Controlling for socioeconomic status, the "black disadvantage is .15 points per month during kindergarten and .19 points per month during first grade, but during

[1]Many Americans have mixed African and European ancestry, amongst other ethnic backgrounds, and can consequently identify as Black, mixed race, or some other race/ethnicity such as White or Hispanic (Bryc, Durand, Macpherson, Reich, & Mountain, 2015). Who has been seen or able to identify as White has varied significantly across time and place in the United States (Haney López, 2006; Omi & Winant, 2014).

summer blacks have a (nonsignificant) *advantage* of .13 points per month" (italics added; p. 624). In other words, the White-Black racial gap in achievement was large during the school year but not statistically significant (and reversed in direction) during the summer. In summary, whereas schooling helps low-socioeconomic-status children make gains in relation to their higher-socioeconomic-status peers, schooling aggravates the gap between Black and non-Hispanic White children. This suggests that racial differences in academic achievement, and therefore educational attainment, are in large part due to differences in the schools that White and Black children attend and/or within-school differences in how White and Black children are treated by educators.

Due to their parents' lower income and wealth and discrimination in the housing market, African American children are more likely to grow up in poor and under-resourced neighborhoods. Children who live in low-socioeconomic-status households are more likely to attend schools with peers who have academic, attention, and behavior problems (Duncan & Magnuson, 2011). These higher concentrations of children with difficulties may make teaching academic content more difficult. Additionally, living in a poor neighborhood often means attending a poorer school. Although schools with high concentrations of low-income students arguably need greater resources than schools serving more advantaged students, they frequently have fewer resources (e.g., Condron & Roscigno, 2003). These financial inequities may worsen socioeconomic gaps in educational achievement.[2] Controlling for prior achievement, schools that spend more on instruction (including teacher salaries) and operations and maintenance have higher academic achievement (Condron & Roscigno, 2003).

In addition to facing financial disadvantages at school, African American students are also more likely to experience instructional barriers than White students. Both experimental and non-experimental research has shown that teachers' expectations are lower for African American children (Tenenbaum & Ruck, 2007). Qualitative research conducted within the United States suggests that lower expectations may lead to reduced achievement through less rigorous instruction and reduced student motivation (Weinstein, 2002). Research outside the United States has found that ethnic-minority children make less growth when they are assigned to teachers who have higher levels of implicit bias (van den Bergh, Denessen, Hornstra, Voeten, & Holland, 2010). Even when bias is not present, African American students tend to be at an educational disadvantage compared to White students with respect to observed teacher qualifications.

African American children frequently attend schools that are highly racially segregated (Orfield, Ee, Frankenberg, & Siegel-Hawley, 2016). Because of segregation, teacher sorting in response to salaries and student characteristics, and principal assignment of teachers, Black children are more likely to be taught by an inexperienced teacher than White children (Clotfelter, Ladd, & Vigdor, 2005; Jackson, 2009).[3] These differences in teacher quality

[2]Some past research that has not found a relationship between spending and academic achievement is limited by its use of district-level data; school spending can vary widely within districts (Condron & Roscigno, 2003).

[3]Clotfelter et al. (2005) found that Black children in North Carolina were more likely to be assigned an inexperienced teacher than White children across the state, within a district, and within a school. Jackson (2009) used the sudden end of busing in Charlotte-Mecklenberg to demonstrate that the relationship between teacher experience and the student race is not simply due to residential patterns.

are not merely due to social class. In the analysis of a national data set, Condron (2009) found that Black children, no matter their social class, were less likely to be taught by a certified teacher than White children; at the same time, social class itself was not related to the chances of being taught by a certified teacher. Even when attending integrated schools, African American children may experience so-called second-generation segregation that channels them into less rigorous classes and White children into rigorous ones, regardless of their objective achievement levels (Lewis & Diamond, 2015; Mickelson, 2015).

In summary, a host of intertwined family- and school-level factors likely contribute racial differences in academic achievement and educational attainment in the United States. The purpose of this dissertation is to not demonstrate the relative strength of these hypotheses. Instead, I provide an in-depth examination of the potential influence of exclusionary school discipline on students' chances of attending college.

2.2 Financial Returns to a College Education

Racial gaps in educational attainment are concerning because of the benefits that additional years of education confer. Although it has been assumed by some that the estimated effects of education are upwardly biased (i.e., that higher education only appears to have a positive effect because more able individuals pursue a higher education), some research suggests that there is negative self-selection into college education (i.e., that those who attend college benefit less from it than others would). Unobserved factors like cognitive ability and self-regulation skills may affect both how much education individuals obtain and how successful they are professionally, but it appears that individuals who are less likely to attend college benefit most from higher education. For example, in a national study, male college graduates ($n = 1,265$) in the lowest propensity score stratum (i.e., those who were least likely to attend) earned 30% more than non-college graduates in this stratum; for those in the highest stratum, college graduates earned only 10% more than non-graduates (Brand & Xie, 2010).[4] This pattern of findings held for two cohorts of both men and women. Individuals with a high propensity to attend college have a socioeconomic background that gives them greater chances of professional success, whether or not they attend college. Individuals with a low propensity to attend college, on the other hand, have fewer opportunities to obtain a well-paying job if they do not obtain the credentials, training, and social networks that a college education provides.

Although earning an associate's degree confers benefits, the returns to a four-year bachelor's degree are higher (Bureau of Labor Statistics, 2016). For example, in 2015, adults with a bachelor's degree who were working full-time earned approximately $1,100 per week while adults with an associate's degree who were working full-time earned approximately $800 per

[4]Like other observational studies, including this dissertation, this research relies upon the assumption that important covariates have been observed. It is possible that college graduates who had a low propensity to attend college may be characterized some important unobserved variable, such as ongoing support from a mentor, that allowed them to overcome disadvantage in terms of college enrollment. If such variables also affect employment and earnings, these findings may not hold.

week. Earning a bachelor's degree is also a necessary step on the path to earning a master's or professional degree; adults with a master's degree earned $1,300 per week, on average, and adults with a professional degree earned approximately $1,700. In 2015, approximately 4% of adults with an associate's degree were unemployed compared to 3% of adults with a bachelor's degree. However, it should be noted that comparisons between individuals who initially enter two- and four year-colleges could be misleading. Approximately 33% of community college students transfer to a four-year college, and approximately 14% of students who begin their post-secondary education in a community college ultimately earn a bachelor's degree (Jenkins & Fink, 2016). Therefore, I examined whether my results changed when I adjusted the outcome to be matriculation at a two- or four-year college rather than a four-year college alone.

2.3 Deciding to Attend College

Why do certain students decide to apply to and attend a four-year college? Students are thought to progress through three stages of college selection (Hossler & Gallagher, 1987; Perna, 2006). First, students become interested in attending college between seventh and tenth grade. Second, students and their parents gather information about colleges and the application process between tenth and twelfth grade. Students usually apply to colleges during the fall of twelfth grade. Third, students who have been admitted to college choose to enroll in a particular school. Both individual- and school-level factors matter in this process, and integrated approaches to understanding the process draw upon both economic and sociological theories.

According to traditional economic theories of educational attainment, individuals choose to pursue a college education if the associated benefits outweigh the costs. For example, Becker (1993) asserts that "human capital investments tend to respond rationally to benefits and costs" (p. 18). Empirical research, however, suggests that college attendance decisions are not as rational as originally proposed by economists. For example, a mere six dollars in savings (due to a change in the ACT policy for sending score reports to colleges) led to substantial differences in college admissions outcomes for low-income students (Pallais, 2015). In response to the change, low- and high-income students sent more score reports and applied to more colleges, and low-income students attended more selective colleges. As the returns to attending a more selective college far exceed six dollars,[5] this suggests that high school students do not act as rationally as proposed by economic theories of college attendance.

Additionally, some economic theories neglect the fact that many individuals do not actively deliberate about whether to attend college. Sociological approaches, on the other hand, to college enrollment incorporate sociocultural and institutional factors, in addition to purely financial ones, and can address this lack of active deliberation. Although the vast majority of

[5]In addition to perhaps conferring higher wages through conferring a more prestigious credential or providing better training, students who attend more selective colleges are more likely to graduate than similar students who attend less selective ones (Alon & Tienda, 2005).

parents – regardless of their own education level (e.g., Carter, 2005; Hayes, 2011) – aspire for their children to attend college, students whose parents have attended college may be more likely to grow up assuming that they too will attend college. For many second- or third-generation college students who grow up in high socioeconomic status households, college attendance is a foregone conclusion. For example, in a study of Texas high school sophomores ($N = 12,522$), students whose parents had completed college were approximately 20% more likely to report that they had always planned to attend college than students whose parents had only completed high school (Grodsky & Riegle-Crumb, 2010). These students' sense of confidence may facilitate their progression through the college selection process (Bourdieu & Passeron, 1964; Grodsky & Riegle-Crumb, 2010), and they can draw up their parents' first-hand knowledge of the college application process.

On the other hand, students with low socioeconomic status may lack the resources needed to apply even if they wish to attend. Therefore, some students who aspire and are qualified to attend college do not apply or are accepted but do not enroll. In a study of students enrolled in Chicago public schools, 83% of students reported that they hoped to earn a bachelor's degree, but only 41% of students actually enrolled in a four-year college (Roderick, Coca, & Nagaoka, 2011). This gap was not merely due to students' overoptimistic aspirations. A gap between aspirations and enrollment was seen even for students who were qualified to attend a very selective four-year college (based on GPA, ACT score, and participation in advanced coursework) and reported that they planned to enroll in a post-secondary institution. Of these students, 10% did not enroll in any college and only 38% in a four-year college that was very selective. Across all students who aspired to attend college, the largest drop-off occurred at the application stage rather than admissions (i.e., a greater proportion of students failed to apply rather than were not admitted). This suggests that students either misjudge the utility of enrolling in a college or their chances of being admitted or cannot effectively surmount the obstacles they encounter during the application process.[6]

This gap may arise in part due to inaccurate or inadequate information. For example, not all parents can estimate college costs, and those who do tend to believe college is more expensive than it actually is (Grodsky & Jones, 2007). It also appears that some of the gap is due to barriers such as applying for financial aid. For example, young people in families that were randomly assigned to receive information and help completing the Free Application for Federal Student Aid (FAFSA) form were much more likely to enroll in college and complete two years of college than those who were not assigned to the intervention (Bettinger, Terry Long, Oreopoulos, & Sanbonmatsu, 2012). Information about the FAFSA and college costs alone did not improve individuals' outcomes compared to the control condition. Individuals in the experimental group still had to independently apply to college and in most instances were required to mail in the FAFSA paperwork themselves. These findings suggest that social and cultural capital, which help individuals navigate the admissions process, and not just financial costs and benefits of attendance, help determine whether adolescents attend college.

Regardless of which theoretical model best represents the process students go through when

[6]It is also possible that their measure of preparedness does not capture some other characteristic, such as a history of suspension, that would affect students' chances of being admitted.

they apply to and enroll in college, it is clear that students' academic preparation and socioeconomic status affect their college admissions outcomes. Students who attend schools with less qualified teachers and less rigorous classes will be less prepared to take college admissions tests and be seen as less desirable applicants by college admissions boards. Academic preparedness is also related to individual household socioeconomic status. For example, household socioeconomic status may affect students' school quality through the financial means to attend a private school, the ability to indirectly choose superior public schools through residential choice, or the skill needed to navigate explicit school-choice systems. Additionally, higher socioeconomic status can directly impact school quality through property-tax-based school funding mechanisms. Additionally, parents with higher levels of income and education are better able to afford private tutoring and ACT/SAT preparation classes (Buchmann, Condron, & Roscigno, 2010).

In addition to socioeconomic status and academic achievement, school policies affect adolescents' decision to apply to college and their success in doing so. For example, high schools may encourage students to visit college campuses, help students with applications for admission and financial aid, communicate directly with college representatives for students, and communicate with parents about their children's college choices (Hill, 2008). Interestingly, however, Hill (2008) found that attending a supportive school increased students' chances of attending a four-year college rather than a two-year college, but did not increase students chances of attending a four-year college compared to not pursuing a post-secondary education. This research was conducted with a nationally representative sample of students who were enrolled in tenth grade in 1990 and may not generalize to more recent times. Research in Chicago public schools found that, controlling for students' background characteristics, students were more likely to plan to attend, apply to, and be accepted by a four-year college if they attended a school with a higher teacher-reported "college-going climate" (Roderick et al., 2011). Students were also more likely to enroll in a college that was an appropriate match, in terms of selectivity (Roderick et al., 2011).

Additionally, school and individual factors may interact: School context may matter more for low-income students. Students whose parents have not attended college have access to less information about colleges and the application process through their social networks than the children of college graduates. Therefore, this first group of students is more dependent on their schools for such information. Research conducted by Carrell and Sacerdote (2013) in New Hampshire high schools supports the hypothesis that non-familial help is especially important for students on the margin of applying to and attending college. Female students assigned to a mentor, who helped the students complete their applications,[7] were much more likely to enroll in college than those given a financial incentive or information alone. Most of the participating students already knew where they wanted to apply to college and used the mentoring time to complete the college and financial aid applications, suggesting that logistical barriers in applying were what mattered. Moreover, the mentoring program had a smaller (or no) effect for students who reported they had help from their parents or

[7]The help provided included support in registering for and paying for the ACT or SAT (if students had not already taken it), completing application forms, writing application essays, paying for application fees, and completing the FAFSA.

teachers in applying for college. Although low-income, first-generation students may need the most support during the college application process, the schools that offer the most supportive policies tend be those who educate students with higher socioeconomic status (Hill, 2008).

In summary, both school characteristics and students' individual characteristics are thought to influence whether a student applies to college and whether he or she is accepted. Research suggests that low-socioeconomic-status and racial-minority students' college attendance is hindered by obstacles on the path to admission, such as obtaining applications, taking standardized tests, and applying for financial aid. In addition to being related to college attendance, students' race/ethnicity, socioeconomic status and academic achievement are also related to their risk of being disciplined. Therefore, any study of the effect of discipline on college matriculation must control for these factors. Otherwise, baseline differences between suspended and non-suspended students may lead to spurious associations between suspension and college attendance. In this dissertation, I analyze how suspension affects adolescents' chances of matriculating at a four-year college after controlling for such differences. As I discuss below, the stigma of suspension may negatively affect adolescents' peer networks, teachers' beliefs, and chances of being admitted through signalling to admission committees.

Chapter 3

Exclusionary Discipline

Exclusionary discipline is a common consequence for misbehavior in American schools. Such discipline can range from office disciplinary referrals, where a student is sent to the office by his or her teacher and misses just 20 minutes of class time, to expulsion, where a student is permanently excluded from school for the remainder of the academic year. Depending on the circumstances, expelled students may or may not receive educational services during their expulsion (U.S. Department of Education, n.d.). In between these two extremes lies suspension. In-school suspension commonly refers to a student being removed from his or her classes for at least half a day and being placed somewhere else on school property under the direct supervision of school personnel (U.S. Department of Education, n.d.).[1] Out-of-school suspension typically refers to a student being excluded from school for at least one day (U.S. Department of Education, n.d.). Although many students may be suspended for only a day or two, suspensions can last for weeks or even months of school (Dominus, 2016). In a study of suspensions in Virginia high schools, 5% lasted between 11 and 364 days (T. Lee, Cornell, Gregory, & Fan, 2011). The 1974 U.S. Supreme Court case *Goss v. Lopez* established that students who are facing suspension must be told what the offense(s) they are being charged with and given an opportunity to defend themselves. In practice, however, these legal rights may be disregarded (Davis, Levine, & Part, 2015; Kupchik, 2010). Even when students are granted an informal hearing, young students may not be able to adequately defend themselves against an adult teacher or administrator who has greater status and power in the school setting. As there is no fundamental right to an education under the current interpretation of the United States Constitution (Kim, Losen, & Hewitt, 2010), schools are not obligated under federal law to help students make up missed instruction, assignments, or exams during suspensions. In fact, some school districts expressly deny students the opportunity to make up missed work during suspensions (e.g., Columbia School District, 2012). In states where students do have a fundamental right to an education, a student who misbehaves may be considered to have waived that right (Kim et al., 2010).

Although the use of exclusionary discipline may be necessary in extreme cases (e.g., when a student poses a physical threat to other students or educators), most instances of exclusionary

[1] Some schools may also refer to shorter removals as in-school suspensions (Kupchik, 2010).

discipline are imposed for non-serious reasons. Qualitative research conducted in a high school suggests that some suspensions are effectively random (Vavrus & Cole, 2002). Vavrus and Cole (2002) found that teachers sometimes made referrals (which resulted in suspensions) for mildly impertinent behaviors that they ignored or tacitly accepted on other days. In a seminal, national study of suspensions during the 1972–1973 school year, the Children's Defense Fund (1975) found that less than 3% of suspensions were for criminal behavior such as the destruction of property or use of illegal drugs. In a more recent study of the reasons for suspension in the largest school district in Florida, "disobedience/insubordination" was the most common reason for suspension (Raffaele Mendez & Knoff, 2003). Only 3% of suspensions were for battery, and less than 1% of suspensions were for possession of weapons or narcotics. Out-of-school suspensions for defiance can be for relatively minor acts such as cursing (Kupchik, 2010, pp. 94-95, 123). Students are sometimes suspended for behavior that poses no clear harm to others, such as falling asleep in class (Kupchik, 2010, p. 207). Gregory and Weinstein (2008) documented students being suspended for chewing gum, being found in "suspicious circumstances," running in the hall, and for simply possessing a marker that presumably might have led to or been connected to vandalism (p. 472). In a study of North Carolina students who had committed just one offense, students were routinely suspended for minor misbehavior such as having a cell phone, violating the dress code, and engaging in a public display of affection, especially if they were Black (D. J. Losen, 2011). Learned (2016) and Kupchik (2010) also report instances where students were suspended for using their cell phones. In summary, students are frequently suspended for behaviors that do not put themselves or others at risk, even though exclusionary discipline has been framed in the past and present as a method of increasing or preserving school safety.

3.1 History of School Discipline and Suspensions in the United States

Much of what is known about suspension dates back to seminal research conducted by the Children's Defense Fund in the 1970s and their analysis of the Office of Civil Rights' data collection on suspensions, which began in 1972 (Children's Defense Fund, 1975). There are currently no thorough historical accounts of the practice of school suspension before the 1970s, although some information exists about general approaches to discipline (Kafka, 2011).

During the time period preceding universal compulsory education in the United States, many schools used corporal punishment (e.g., beating) and other physical punishment (e.g., standing for long periods of time) to bring about students' submission. Some schools, such as those promoted by Horace Mann, attempted to bring about a disciplined environment through promoting a genuine environment of mutual respect and regard. Kafka (2011) only mentions the exclusion of students — locking them in a dark place — as a way to bring about obedience through fear, not to exclude them from the classroom per se, during this time period (p. 24). However, Seeley (1901) argued that when faced with an "incorrigible pupil," "sometimes the teacher owes it to the school to suspend a child" despite the fact

that this "may throw the boy upon the street, may send him to destruction" (p. 82). Seeley (1901) also noted that suspensions were decided upon by teachers rather than administrators (p. 83). In summary, it is unclear how frequently exclusion from the classroom or school was used as a punishment in the 18th and 19th centuries and for how long such suspensions lasted when they did occur.

From the late 19th century through World War II, some schools sought to bring about order through a democratic process and a positive focus on individual development (Kafka, 2011). However, these approaches were not necessarily widespread, and many schools continued to use corporal punishment. Some "troubled" students were transferred to separate schools, as occurs with expulsion today (Kafka, 2011, p. 41). It also appears that school suspensions occurred, although not necessarily frequently. In a historical study of psychotherapy for children aged six to sixteen years living on the East Coast, the author reported that three out of 50 children (6%) had been suspended from school (Glassman, 1943).

During the 1950s, members of the public became increasingly concerned about a perceived increase in school violence and pervasive juvenile delinquency, which did not necessarily reflect reality (Kafka, 2011), as in the 1980s and 1990s (Hirschfield, 2008; Kupchik, 2010). It appears that "problem" youth were sometimes transferred to separate educational settings (Kafka, 2011, p. 55), but it is unclear whether temporary in- or out-of-school suspensions were used frequently in schools across the United States at this time. In the Los Angeles unified school district, which was beset by overcrowding and other stress to the school system, suspensions were reportedly "virtually nonexistent" in the 1950s, although approximately 1% of middle and high students were placed in an alternative setting (Kafka, 2011, p. 61, 141). On the other hand, I find reference to school suspension in a high school handbook published and reportedly distributed to students in Massachusetts in 1957: "A pupil will be suspended from school for any willful disobedience, open and persistent defiance of the authority of the teacher, habitual profanity or vulgarity, or smoking on school premises (Czrnawski, 1954, p. 35). No explanation of what constituted a suspension was given. Czrnawski (1954) reported that "detention" was discussed in 9 out of the 35 school handbooks she reviewed, but the existence and frequency of a section devoted to suspension in these other handbooks was not noted.

Regardless of the prevalence of exclusionary discipline in the past, it is clear that many school districts were suspending and expelling students when more schools began to racially integrate following the *Brown v. Board of Education*[2] rulings and the Civil Rights Movement. Researchers from the Children's Defense Fund analyzed data from the Census and Office of Civil Rights and interviewed more than 6,500 families and 300 school officials in nine states and the District of Columbia. During the 1972–1973 school year, approximately 4% of all students and 8% of high school students were suspended at least once (Children's Defense Fund, 1975). These percentages are likely underestimates as they did not include students who dropped out from school (who are more likely to be suspended than students who do not drop out) or students who were sent home for "cooling-off periods" (p. 10). Some districts also appear to have systematically under-reported the number of suspensions

[2]Although Brown mandated that schools desegregate with "all deliberate speed", racial integration did not occur in some districts until the late 1970s (e.g., Kafka, 2011).

that occurred. Students who were suspended were excluded from school for four days on average, although some districts excluded students for long periods of time. For example, in Baltimore, Maryland, suspended students missed 35 days of school, on average (Children's Defense Fund, 1975, p. 56).

As is the case today, students were frequently suspended in the past for behaviors that were neither violent nor harmful to other students' learning. The majority of suspensions documented by the Children's Defense Fund (1974, 1975) were for non-violent offenses (e.g., truancy). In some districts and schools, more than 50% of suspensions were given for truancy or tardiness. In their survey of students, approximately 36% of suspensions were for fighting with another student, 2% were for destruction of property, 1% were for fighting with a teacher, and less than 1% were for illegal drugs or alcohol (Children's Defense Fund, 1975, p. 38). In some instances, children's offenses were minor acts of disobedience that reflected an underlying physical problem with the student. For example, a 11-year-old girl was suspended for three days after she left class to use the bathroom when her teacher would not give her permission to do so. In another instance, a student was suspended twice (for one week and one month) for eating in class; it was later discovered that she had a brain tumor affecting her hypothalamus, which made her feel hungry constantly.

The Children's Defense Fund (1975) described a popular belief that suspensions were a last-resort punishment that dissuaded students from being violent or disruptive; this was not accurate in the 1970s and is not accurate today. Interestingly, however, none of the administrators interviewed by the Children's Defense Fund (1975) stated that suspensions were a way to help children. In the words of one interviewee, "Suspensions serve no purpose at all only that it might worsen the problem" (p. 15). At best, suspensions were seen as a way to temporarily free educators from dealing with troublesome students and ostensibly "to get parents in" (p. 16). However, many parents were never notified of their child's suspension and a third did not attend a school conference. In some instances where parents were required to attend a conference but could not because of lack of child care or health issues, for example, children remained out of school for weeks.

In a more recent qualitative study, some disciplinarians also described suspension as a way to involve parents and even punish parents themselves (e.g., by making them miss work to attend a conference) (Bowditch, 1993). It is unclear if these beliefs are true for some disciplinarians today or if they were ever prominent. This research was conducted in the mid-1980s at an inner-city school whose student body was entirely composed of African American students. Many of these students lived in poverty, and most had low academic achievement. Their parents mostly lacked the capital necessary to challenge school authority.

Although some school administrators acknowledged that suspensions were not likely effective or served mainly as a way to get parents' attention, others believed that suspensions would decrease misbehavior and therefore be beneficial to students (Sanders & Mendez, 1981). Consequently, zero-tolerance policies, which mandate automatic suspension or expulsion for certain categories of offenses, no matter the circumstances, became more common in the decades following the Children's Defense Fund's seminal study. This occurred in spite of the *Goss v. Lopez* ruling in 1975 that students were entitled to some protections before being suspended and that suspensions were stigmatizing, in addition to depriving students

of instruction (Kim et al., 2010). Some school districts such as Los Angeles instituted zero-tolerance policies for drugs and weapons possession in the 1970s (Kafka, 2011), and nationwide federal policies followed in the 1990s. The Gun-Free Schools Act, which was enacted in 1994, required schools receiving federal funding to expel for one year any student found with a weapon. In response to this legislation, students were expelled for having guns but also items such as pocket knives that were not intended to be used as weapons against another person (American Psychological Association Zero Tolerance Task Force, 2008).

Although these laws did not mandate suspensions for other offenses, it is thought that they contributed to the extension of a zero-tolerance framework to non-dangerous behaviors such as using a cell phone (American Psychological Association Zero Tolerance Task Force, 2008; Kim et al., 2010). As zero-tolerance policies expanded, many schools also began to hire police officers to work on site (Kupchik, 2010). While some teachers and students report that such officers make schools feel safer, their presence can lead to the escalation of minor incidents into ones that end in suspension and/or arrest (Kupchik, 2010). Mirroring this expansion of zero-tolerance policies and police presence on school grounds, it appears that high school suspensions have increased since the 1970s, especially for Black students (D. Losen, Hodson, Keith, II, Morrison, & Belway, 2015). However, these trends should be interpreted with caution, as the aforementioned under-reporting makes comparisons difficult. During the 1988–1989 school year, 10% of Black students in kindergarten through Grade 12 were suspended, compared to 6% in 1972–1973. During the 1988–1989 school year, 5% of White students were suspended, compared to 3% in 1972–1973. During the 2011–2012 year, White students' suspension rate remained at 5%, and Black students' suspension rate further increased to 16%.

In summary, temporarily removing students from school has been used as a punishment since at least the beginning of the 20th century. However, systematic data collection of suspensions in public schools only began in the 1970s, and it remains unclear how frequently suspensions were used before this time. Suspensions have apparently increased since the 1970s, although concerns about under-reporting during the initial rounds of the Office of Civil Rights data collection make the trend for White students less clear. Some educators and policymakers have argued that suspensions are necessary to preserve school order or safety, and have used this position to establish through zero-tolerance policies in schools. However, there is little evidence that such policies actually make schools safer, and suspensions have frequently been used to punish students' for behaviors that do not jeopardize other students' learning or safety.

3.2 Disproportionate Discipline of Black Students

Studies at the school, district, state and national level have consistently found that African American students are more likely to be subjected to exclusionary discipline than White or Asian American students (e.g., Fabelo et al., 2011; S. B. Horner, Fireman, & Wang, 2010; Skiba et al., 2011; Sullivan et al., 2013). Based on the analysis of Office of Civil Rights data, African American students were approximately twice as likely as White students to be

suspended in the early 1970s (Children's Defense Fund, 1975, p. 9). Survey data collected by the Children's Defense Fund (1974) during this same period suggests that the gap may have been worse for high school students, with African American students being three times as likely to be suspended as White students. Since this time, racial gaps in discipline appeared to have stagnated or worsened.

Students who identified as African American or Black made up just 16% of the student population during the 2011–2012 academic year but 32% of students who received an in-school suspension, 33% of students who received one out-of-school suspension, and 42% of students who received multiple school suspensions (U.S. Department of Education Office for Civil Rights [OCR], 2014). White students made up 51% of enrollment but only 40% of students who received an in-school suspension, 36% of students who received one out-of-school suspension, and 31% of students who received multiple out-of-school suspensions (U.S. Department of Education Office for Civil Rights [OCR], 2014).

These differences in rates of discipline cannot be wholly accounted for by differences in behavior. After conditioning on having committed the same offense, Black students are still more likely to be suspended than White students (D. J. Losen, 2011; Nicholson-Crotty, Birchmeier, & Valentine, 2009). For example, in analysis of Missouri schools, Nicholson-Crotty et al. (2009) found that 55% of Black students disciplined for smoking were given an out-of-school suspension compared to 37% of White students disciplined for the same offense, and 95% of Black students disciplined for a weapons violation were given an out-of-school suspension compared to 85% of White students disciplined for the same offense. Harsher consequences for Black students for these offenses do not seem to reflect a rational response to greater rates of misbehavior or recidivism, as Black and White students report similar rates of misbehavior with respect to these infractions. In a national survey of high school students, 5.7% of White students reported carrying a weapon at least one day during the 30 days preceding the survey compared to 3.9% of Black students (Zhang, Musu-Gillette, & Oudekerk, 2016). Individual tobacco use was not reported, but similar percentages of Black (18.6%) and White (20.4%) students reported that illicit drugs (including cigarettes) were made available to them at school (Zhang et al., 2016).

The disproportionate discipline of Black students has also been documented for offenses that do not put students themselves or others at risk. For example, in a study of North Carolina schools, Black students referred for a first-time offense were approximately twice as likely to be suspended for cell phone and dress code violations, being disruptive, and engaging in a public display of affection than White students (D. J. Losen, 2011). In a national sample of schools using school-wide positive behavior supports,[3] Black children who were referred were more likely to receive an out-of-school suspension or be expelled than White children who were referred, after controlling for the type of infraction.

Experimental research suggests that teachers are more likely to believe Black students will misbehave in the future and to see Black students as "troublemakers," compared to White students who are otherwise identical (Kunesh & Noltemeyer, 2015; Okonofua & Eberhardt, 2015); these perceptions may affect teachers' interpretations of ambiguous student behav-

[3]This approach is thought to potentially *reduce* bias by standardizing consequences for behavior.

ior and their reactions to actual instances of misbehavior (Kupchik, 2010, pp. 176–178). Okonofua and Eberhardt (2015) found that practicing teachers ($N = 57$) who reviewed a school disciplinary record with a stereotypically African American name reported feeling more troubled about a second infraction and recommended harsher punishment than those who reviewed a record with a stereotypically European American name. In a second study of practicing teachers ($N = 191$), teachers assigned to the African American condition were more likely to report that the student was a troublemaker, that his behavior was indicative of a pattern, and that they would suspend him in the future than those assigned to the European American condition (Okonofua & Eberhardt, 2015). However, this research study used African American names (Darnell or Deshawn) that may be more associated with lower socioeconomic status than the stereotypical White American names employed (Greg or Jake). To lessen the potential influence of socioeconomic status, Kunesh and Noltemeyer (2015) assigned pre-service teachers ($N = 98$) to a vignette featuring either a stereotypically African American name (Darius) or a stereotypically European American name frequently given by parents with low socioeconomic status (Cody). Pre-service teachers assigned to a vignette featuring an African American student reported that the problematic behavior was more likely recur in the future than those assigned to a vignette featuring the European American student. Although this approach may have lessened the confounding influence of socioeconomic status, the generalizability of these results is still unclear. Many Black Americans do not have stereotypically African American names; therefore, this experimental paradigm may be capturing a specific type of racial bias that does not extend to all Black people living in the United States.

Despite experimental evidence and intensive qualitative evidence that demonstrates that children of different racial backgrounds are treated differently by the same teachers and administrators, some have argued that racial disparities in discipline are due to differences in disciplinary policies across schools rather than racial bias or discrimination within schools.

In a study of approximately 1,000 North Carolina schools, Kinsler (2011) compared a) whether a recorded, referred offense resulted in a suspension and b) the length of suspensions for Black and White students. While racial differences in whether students are suspended and differences in length of suspensions are indeed evidence of bias, a lack of differences does not imply that discrimination is absent. Being labeled and recorded as having offended is not the same thing as actually offending. Many students commit offenses that are ignored, and subjective behavior that does not necessarily constitute an offense may be recorded as one. Therefore, racial bias would still exist if Black students were more likely than White students to be referred for the same behavior even if consequences after referral were identical. For example, if Black students are referred eight out of ten times for a certain behavior and White students are referred two out of ten times for that same behavior, racial bias exists even if all referred students are suspended.

Additionally, Kinsler (2011) defined racial bias as "differential punishment for otherwise identical black and white students who commit the same offense," and he argued that students who attended different schools could not be compared. Given the degree of systematic racial segregation in the United States (Reardon & Owens, 2014), this definition is problematic. Although African Americans make up only approximately 14% of the population in

the United States (Rastogi, Johnson, Hoeffel, & Drewery, 2011), millions of African American children attend schools where minorities make up more than 90% of the student body (Orfield et al., 2016). Although there is a degree of randomness in administrator assignment and administrator beliefs about discipline, systematically higher referrals and harsher punishment in high-minority schools is indeed evidence of structural, racial bias in school discipline.

In this study, Kinsler (2011) found that students who were referred for violating school rules were more likely to be suspended and receive longer suspensions if they were Black. For example, Black elementary school students were 10% more likely than White elementary school students to be suspended after having received a discipline referral. However, Kinsler concluded that racial bias did not exist because the association between disciplinary consequences and student race disappeared after including fixed effects for schools in the model.[4] If Black students systematically attend schools with harsher punishment regimes, this is evidence of racial bias — not a lack of it. Indeed, Skiba et al. (2014) found that students were more likely to receive out-of-school suspensions than in-school suspensions after having been referred as the proportion of Black students in a school increased; they controlled for student race and the type of offense students were referred for.

Kinsler (2011) also concluded that racial discrimination did not exist because there was "little evidence" that Black elementary school students were treated differently by Black teachers and racial disproportionality existed in schools headed by Black principals[5] (p. 1130). This approach is not an adequate test of racial bias, and this finding is not evidence of lack of racial discrimination. Messages and stereotypes about Black criminality are omnipresent in American society. Consequently, African Americans can also experience implicit bias against African Americans. Based on the analysis of Implicit Association Test data, people who identified as Black more easily formed associations between other Black people and weapons and between White people and harmless objects than the reverse; this was also true of people who identified with other ethnic/racial backgrounds (Nosek et al., 2007). This finding suggests that Black educators may harbor some of same implicit biases against African American students as White educators. Additionally, knowing that African American students will face discrimination from many members of society, African American educators may treat these students firmly in an attempt to prepare them for the future. In particular, Black male teachers may be positioned as disciplinarians and be expected to harshly discipline their African American students, sometimes against their own wishes (Brockenbrough, 2015).

Finally, Kinsler claims that "the most obvious explanation" for disparities in referrals "is simply higher levels of disruptive behavior among black students" (p. 1379). On the contrary,

[4]Kinsler (2011) also controls for students' disciplinary records. As race itself may be a primary reason for past *recorded* offenses, this variable should arguably not be included. He also noted that the effect of race remained significant after only controlling for observed school-level variables, but he did not report these findings in detail.

[5]Kinsler (2011) examined the effect of principal race in one third of the schools in his sample. Although principals play a large role in establishing a school's discipline policy (Skiba et al., 2014), assistant or vice principals, rather than principals, are frequently in charge of determining specific consequences for students referred to the office by teachers.

both past and recent research has found that Black and White students report similar levels of rule breaking (e.g., McCarthy & Hoge, 1987; Kann et al., 2014; Zhang et al., 2016). In a study of two high schools in the American southwest and two high schools in a mid-Atlantic state, Black students were more likely to be disciplined at school than White students, even after controlling for self-reported delinquency, history of skipping classes, and criminal history (Kupchik, 2010); the gender gap in discipline, however, disappeared after controlling for the aforementioned variables. The persistence of the racial gap after controlling for proxies for behavior suggests that that the disproportionate discipline of Black students is due to factors of than differential rates of misbehavior. McCarthy and Hoge (1987) investigated whether systematic under-reporting of misbehavior amongst Black students was present, and they found no evidence for this hypothesis. While further studies of student behavior may elucidate the processes at play in the disproportionate discipline, far more evidence supports the hypothesis of the role of differential treatment of African American students than that of differences in student behavior.

Setting aside these problematic assumptions and interpretations, Kinsler's finding that the effect of race disappears after including school fixed effects is interesting, especially in light of other within-school research that finds Black students are treated differently (e.g., Lewis & Diamond, 2015) and research that finds that the association between race and discipline remains statistically significant after accounting for unobserved school variation using mixed effects (e.g., Sullivan et al., 2013) and fixed effects in other geographical areas (e.g., Morris & Perry, 2016). These unexpected results should prompt researchers to do more within-school research across the United States in order to determine whether these findings are specific to the North Carolina context or time period studied or primarily a result of Kinsler's methodological choices. Not all schools punish African American students to a disproportionate degree, and closer analyses of schools with low and high levels of racial disproportionality could elucidate the processes which lead to discipline gaps and potential solutions to the problem.

3.3 Consequences of Suspension

Students who are suspended tend to have worse academic outcomes than their non-suspended peers. There are two main explanations for this association: suspended students lose instruction time and are frequently stigmatizated, which may affect other individuals' expectations for such students and students' future beliefs and behaviors. However, many previous attempts to identify the consequences of suspension have been weakened by potential reverse causality and confounding.

Apparent associations between suspension and poor academic outcomes may be due to baseline differences between suspended and non-suspended students. For example, students who qualify for free or reduced-price lunch (and therefore live in households below or close to the poverty line) are more likely to be suspended (Sullivan et al., 2013). Schools may be more reluctant to suspend children whose parents have high socioeconomic status because they believe these parents are more aware of the children's legal rights and more likely to effectively

challenge the schools decisions (Lewis & Diamond, 2015). Additionally, high-socioeconomic-status students themselves can negotiate more effectively with teachers and administrators when they are caught misbehaving, thereby avoiding harsh punishments (Kupchik, 2010). Such students are socialized from a young age to negotiate with authority figures in effective ways (Lareau, 2011). Low-socioeconomic-status and ethnic-minority students' attempts to be assertive in interactions with educators, however, may backfire (Carter, 2005). This relationship between socioeconomic status and discipline is problematic because students who live in low-income households are also less likely to attend a four-year college (Caro et al., 2015; Feliciano & Ashtiani, 2012). It is therefore essential to control for socioeconomic status when studying the relationship between suspension and college attendance.

In addition to socioeconomic status, pre-existing differences in academic achievement may explain the association between suspension and educational outcomes. Historically, children were punished for failing to master academic material in school (Kafka, 2011), and an association between students' academic achievement and their chances of being disciplined persist to this day. Some researchers have hypothesized that low-achieving students are more likely to be disciplined and eventually excluded from school due to pressures related to high-stakes testing (Figlio, 2006; Hirschfield, 2008). Teachers and administrators may consciously or unconsciously want to remove students whom they suspect will score poorly on standardized tests used for school accountability. Conversely, teachers may view high-performing students as "good kids" and treat them more leniently when they misbehave (Kupchik, 2010, p. 174). Bowditch (1993) also found that high-achieving students who misbehaved were treated differently than low-achieving students. Students' "grades altered the meaning of [their] behavior" in administrators' eyes (p. 501).

Low academic skills may also increase the likelihood that students misbehave, putting them at higher risk for suspension. For example, in a study of four high schools, Kupchik (2010) found that "many students misbehave in class because they do not understand the course material; misbehaving distracts attention away from these students' academic failures [and] brings them positive peer attention" (p. 6). In these instances, low academic skills can lead to suspension. Additionally, students in remedial classes may be driven to misbehavior by the lack of engaging work, which they may perceive to be "meaningless" (Learned, 2016, p. 1288); in this case, lack of rigorous instruction would be a source of both lower achievement and "problem" behavior that leads to disciplinary consequences. Finally, it is possible that suspended students have behavior problems that put them at risk for academic failure, and it is not suspension *per se* that worsens academic outcomes. Therefore, it is essential to control for baseline academic achievement as this variable is strongly associated with college attendance (Perna, 2006).

Unfortunately, most of the existing research reviewed by Noltemeyer et al. (2015) did not control for pre-existing differences in academic achievement. Additionally, studies that have controlled for socioeconomic status have often been limited to using coarse measures such as eligibility for free and reduced-price lunch eligibility rather than household income, parental occupation, or parental education (e.g., Arcia, 2006). Therefore, the fact that suspension has been found to be negatively associated with academic achievement and positively associated with drop out may not reflect a causal relationship. However, there are a select number

of peer-reviewed papers and non-peer-reviewed reports that have attempted to control for baseline differences in academic competency.

One of these studies examined the association between suspension and subsequent reading and math achievement of students ($N = 16,248$) attending public middle and high schools in a large school district in Kentucky (Morris & Perry, 2016). They accounted for the influence of socioeconomic status with a dichotomous variable — whether the student qualified for free or reduced-price lunch during that school year year — as well as student gender, race, special education status, and household structure. Although they do not control for baseline math and reading achievement, they use longitudinal data and multilevel models to account for unobserved student heterogeneity that is constant over time and individual differences in growth over time. Compared to themselves at other times, students who were suspended in a given year had statistically significantly lower math and reading achievement, although the differences were small in terms of educational significance. Much larger between-student effects for suspension were found, but these estimated effects of suspension likely remain confounded by baseline differences between suspended and non-suspended students.

In a longitudinal study of 928,940 students, Fabelo et al. (2011) found that 31% of disciplined students (where discipline included both suspension and expulsion) were retained, or held back in a grade, at least once compared to 5% of non-disciplined students. Ten percent of disciplined students dropped out from school compared to only 2% of non-disciplined students. After controlling for past disciplinary action and other potential confounds, they found that students who were disciplined were more likely to be retained in a grade; they do not report the results of a multiple regression analysis of drop out as an outcome. Fabelo et al. (2011) included indicator variables for a) whether students had ever failed the Texas state achievement test and b) whether they failed at least one section of the Texas state achievement test in the last year they took the exam, but they did not control for a more fine-grained measure of academic achievement. Moreover, it is unclear whether one or both of these variables were post-treatment, as the authors note that the variable was based on whether they failed a test "before or *during* [emphasis added] the study period" (Fabelo et al., 2011, p. 89). Controlling for post-treatment variables (i.e., potential consequences of treatment) results in biased estimates. Finally, these data were collected in Texas and may not necessarily generalize to other contexts.

Other studies that control for academic achievement are impacted by data limitations or methodological decisions. For example, Balfanz, Byrnes, and Fox (2014) found that students who were suspended in ninth grade during the 2000-2001 school year were less likely to complete high school or attend a post-secondary school. Although they controlled for students' attendance rate and having failed classes, they were not able to control for gender due the nature of their data set. This is problematic as boys are much more likely to be suspended, more likely to drop out, and less likely to attend college than girls. Additionally, attendance rate and failing a course were potentially post-treatment outcomes in their data set; they controlled for missed school days and courses failed during ninth grade, which frequently would have occurred after having been suspended. This study was conducted in Florida and approximately 60% of students qualified for free or reduced-price lunch.

There also exists some evidence at the school level that suspension negatively impacts aca-

demic outcomes. In a study of Virginia public high schools, T. Lee et al. (2011) found that school suspension rates were positively related to school drop out rates, even after controlling for student body demographics, percent of students who qualified for free or reduced-price lunch, students' aggressive attitudes, and students' belief in school rules. These findings are interesting as they control for a proxy for student misbehavior (Stewart, 2003). However, they do not control for students' middle school achievement.[6]

Despite the weaknesses of much of this research, there are clear theoretical reasons to believe that there is a causal relationship between suspension and other forms of exclusionary discipline and academic outcomes. Students who are suspended miss instruction during the time they are removed from class or the school itself. Systematic studies of the duration of suspensions are rare. However, in a study of Florida ninth graders in 2000-2001, students who were ever suspended were on average suspended twice and for a total duration of seven days (Balfanz et al., 2014). Forty percent of suspended students missed at least five days of school.

As suspended students are frequently academically behind to begin with, they may be especially negatively impacted by a forced absence from class although in- and out-of-school suspensions typically last no more than a few days. Suspended students usually do not receive compensatory instruction. In addition to reducing instructional time, suspensions may directly affect academic performance by preventing students from handing in assignments and taking tests that count toward their grade, depending on state and school district regulations (Kim et al., 2010). Students who are transferred to alternative schools during or following a suspension may be unable to earn a high school diploma (Kim et al., 2010). Instead, students may be required to work toward a GED, which is inadequate for admission to many colleges. In some states, mainstream schools can refuse to accept course credits earned in alternative schools (Kim et al., 2010). Therefore, students who transfer back to the traditional school system may still be at a disadvantage when they resume their work towards a high school diploma.

Suspensions may also impact academic performance indirectly, through motivational channels. If students feel they have been treated unfairly by school staff, they may become less motivated to seek help from the teachers whose referral lead to their suspension. Suspended students may perceive that referring teachers have lower academic expectations and care less about them than non-referring teachers do (Gregory & Weinstein, 2008). By repeatedly suspending students, schools may communicate that they are not committed to their educational success and do not value them as persons (Bowditch, 1993). As one student interviewed by the Children's Defense Fund (1975) explained, "If I got one thing out of those suspensions it was that I was a guy without any worth in the world" (p. 48). These messages may naturally decrease students' academic engagement. If suspensions reduce academic success through reduced opportunities to learn, for example, this may also affect students' motivation. Educational success is implicitly believed by many to be at least in part a function of intelligence in a purportedly meritocratic system. Given that both educational success and intelligence are highly valued by members of society, it has been theorized that academic

[6]It would be inappropriate to control for high school achievement, as this is thought to be affected by suspension.

failure may push young people to reject school and society itself in order to preserve their self-worth (Bourdieu, 1994, pp. 49-50).

Additionally, students who are suspended may come to be labeled implicitly or explicitly as troublemakers by their teachers because of the history of suspension alone (Okonofua & Eberhardt, 2015). These students may then be seen as individuals who are not academically motivated, and this perception may shape teachers' behavior toward such students. For example, these perceptions could inadvertently affect teachers' grading, thereby directly affecting students' academic records. Altered expectations may also influence a teacher's behavior toward a student in more subtle ways, such as by affecting the academic rigor of the material that is assigned, the type and number of questions posed to students, and the amount of encouragement that is offered (Weinstein, 2002). Students who are perceived to be troublemakers and not academically motivated will get less enthusiastic recommendation letters for their college applications.

A similar process of stigmatization and labelling may happen with peers (e.g., Children's Defense Fund, 1975, p. 46). Labeling, secondary deviance, and peer deviancy/delinquency training theories predict that students who are labeled as deviant, such as through being suspended, will be stigmatized, rejected by non-stigmatized peers, exposed to "deviant" peers, and learn more "deviant" behaviors because of this exposure (Bernburg, Krohn, & Rivera, 2006).

Moreover, a history of school suspension may act as a direct signal to college admission boards themselves, as many four-year college application forms include a question about whether students have ever been suspended. In 2006, the Common Application, which is used by hundreds of colleges, began asking students to report whether they had been suspended (or subject to other disciplinary action) (Weissman & NaPier, 2015). School staff are also asked through the Common Application to provide information about whether students have been suspended. This change follows the collection of the data used in this study, but it appears that students were also asked to provide information about their disciplinary history on many individual college admission forms before this change (Gordon, 2007).[7] Weissman and NaPier (2015) found that approximately three quarters of American colleges and universities that participated in their study asked about students' disciplinary records, and almost 90% of these colleges reported using such information when making admissions decisions. Although the estimates reported by Balfanz et al. (2014) surely reflect confounding due to gender, their finding that the movement from zero to one suspension is associated with the biggest drop in academic outcomes suggests that signalling — perhaps to peers, teachers, or college admissions boards — may be an especially important factor.

In summary, there are psychoeducational theoretical reasons to believe that suspension would negatively affect academic achievement, and much correlational research has demonstrated that suspension is associated with poor educational outcomes. However, one recently published article argues that suspension may actually have a positive effect on students' academic achievement.

[7]Some colleges, such as the University of California system, do not use the Common Application and do not ask students to provide information about their disciplinary history.

This article analyzed the test score outcomes of students enrolled in middle schools in three North Carolina school districts during the 2000-2001 school year (Kinsler, 2013). The analysis was based on a model of student and school "behavior"[8] that was composed of three components: "a student utility function over behavior, an achievement production function, and a principal utility function over discipline" (p. 362). Kinsler also used mixture models to classify students as either "'good' [or] "'bad' seeds" (p. 367). Kinsler found that students' own and peers' out-of-school suspensions were associated with students' predicted lower achievement in math, even after controlling for observed student characteristics and purported student "type." He found that a one standard deviation increase in out-of-school suspensions for minor infractions in a school was associated with a 2.1% *decrease* in scores on a math achievement test administered at the end of the school year; a one standard deviation increase in out-of-school suspensions for serious infractions in a school was associated with a 2.0% *decrease* in math achievement. However, because he found that the predicted number of out-of-school suspensions is lower in schools with longer suspension times, he concluded that suspensions deter students from misbehaving. This, of course, is not necessarily the case. He merely found that schools that had lower out-of-school suspension rates tended to suspend students for longer when they were suspended. This may be because schools that suspended students less frequently reserved out-of-school suspensions for more serious instances of misbehavior.

This analysis also relied on several problematic assumptions. Despite abundant educational research that shows teachers and administrators inconsistently enforce rules, e.g., (Kupchik, 2010, pp. 174, 180, 184, 189), Kinsler assumed that schools follow strict, consistent, and unchanging disciplinary policies throughout the school year and that all students face the same consequences for their misbehavior (p. 356). These assumptions are clearly unfounded: "School practices lead to children being disciplined differently ... despite the school's stated goals to the contrary" (Lewis & Diamond, 2015, p. 168). Indeed, even students involved in the same incident frequently receive different punishments (Figlio, 2006). Kinsler also ignored the fact that teachers communicate about students and share information about students they perceive to be "troublemakers" (Kupchik, 2010, pp. 176–178) and that students have files that may follow them throughout their schooling. He instead assumed that students "begin the school year with a clean behavioral slate and are cognizant of the discipline policy" (p. 364). As educators and administrators enforce rules inconsistently, many students are surely not cognizant of their school's discipline policy and cannot correctly anticipate the consequences for their behavior.

Additionally, Kinsler (2013) assumed that having a recorded offense in one's school record was the same as having committed an infraction when in fact these things are not equatable. What he termed an "infraction" and "committing an offense" was actually receiving an out-of-school suspension. As described above, students may offend but not be referred, and students may be referred but not be suspended (e.g., D. J. Losen, 2011; Nicholson-Crotty et al., 2009). In some instances of ambiguous misbehavior such as disrespect, students may not offend, but still be referred and suspended (e.g., Skiba, Michael, Nardo, & Peterson,

[8]As explain below, Kinsler (2013) observed reported reasons for out-of-school suspensions rather than actual instances of misbehavior.

2002).

In summary, most existing research suggests a negative relationship between exclusionary discipline and academic achievement and a positive relationship between exclusionary discipline and dropping out from school. However, the majority of this research has failed to control for baseline differences in academic achievement or is limited in its geographical scope. Given the existence of some research that purports suspension could actually have a positive effect on students' academic achievement (Kinsler, 2013), it is essential for more rigorous analyses to be conducted. I build upon existing findings by using data that was collected across the United States, controlling for an objective measure of academic achievement at baseline, and extending my analysis to college attendance, which is currently an understudied consequence of suspension.

Chapter 4

Methods

The estimated "effects" of independent variables in observational studies frequently do not represent true causal effects. Observational studies are subject to potential selection bias – individuals may self-select into treatment. Unobserved variables may be causally related both to predictor and outcome variables, resulting in confounding and inaccurate estimates. Educators, policymakers, and researchers are generally interested in the *effect* of suspension on academic outcomes, but suspension is not randomly assigned and is therefore not independent of an individual's potential outcomes (i.e., what would happen if the student were or were not suspended).

Randomized experiments allow researchers to estimate causal effects but cannot always be implemented because of ethical or practical concerns. This is especially true in the case of educational settings such as this one. Although technically feasible, it seems unlikely than a school district would approve a randomized controlled trial to allow researchers to study the effects of suspension. Causal inference techniques, such as matching, attempt to improve upon traditional techniques for analyzing observational data. Although matching still cannot produce causal estimates when unobserved confounders are present, it is a promising research technique when the majority of theoretically important determinants of the treatment and outcome are observed. Additionally, the explicit assumptions of the potential outcomes framework can guide the development of appropriate research questions and subsequent analyses.

4.1 Potential Outcomes Framework

Under the potential outcomes framework, the causal effect of a treatment is defined as the difference between an individual's potential outcomes: His or her outcome under treatment is compared to his or her outcome under control (Rubin, 1974). For example, one student may attend college if she is not suspended but not attend if she is suspended. Another student may attended college whether or not she is suspended. Only one of an individual's potential outcomes can be observed, which is the fundamental problem of causal inference

(Holland, 1986).

In this framework, it is also understood that there can be "no causation without manipulation" or induced change (Holland, 1986, p. 959). According to this position, truly *causal* effects can only be estimated when one can at least imagine an experiment in which the treatment were manipulated by an experimenter. Although it would likely be unethical (given the existing correlational evidence suggesting that suspension is harmful), one can imagine an experiment where misbehaving students are either randomly assigned to be suspended or disciplined in another way such as serving detention or repairing their offense (e.g., paying for damaged property).

Through matching, I attempted to infer what a student's outcome would have been if he or she had not been suspended. In other words, I observed suspended individuals' outcomes under treatment $Y_i(1)$, and I estimated their potential outcomes under control $Y_i(0)$ by matching suspended individuals to similar (in terms of potential pre-treatment confounders X_i) yet non-suspended individuals. After stratifying on the selection variables X_i, suspension can be considered as-if randomly assigned (Morgan & Winship, 2015). The mean difference between the treated individuals' outcomes and the matched control individuals' outcomes will then be an unbiased estimator of the sample average treatment effect on the treated (SATT). If treatment assignment is not independent of potential outcomes given covariates X_i, however, causal estimates cannot be recovered. In other words, treatment assignment (whether or not a student is suspended) should be independent of potential outcomes (a student's college matriculation outcomes when she or he is or is not suspended) after controlling for covariates X_i. If this assumption is met, matching can approximate a randomized experiment and the difference in means will represent an estimate of the causal effect. Given that significant variation exists in school disciplinary policies, there is reason to believe that such randomness could be captured after conditioning on observed covariates: "a child's chances of being suspended depend on which district, which school, and which class he or she has the *luck* to land in" (Children's Defense Fund, 1975, p. 15, emphasis added).

In order for the estimated SATT to be meaningful, I must also assume that individuals assigned to treatment could have been assigned to control (Heckman, Ichimura, & Todd, 1998, p. 263):

$$0 < Pr(T_i = 0 \mid X) < 1 \tag{4.1}$$

This assumption is reasonable here. Students who commit an egregious offense necessitating exclusionary discipline are expelled rather than suspended, and these students are not included in the following analysis. In American schools, suspensions are most commonly used for misbehaviors that could otherwise be ignored or responded to in a more lenient way, such as after-school detention (e.g., Raffaele Mendez & Knoff, 2003). Moreover, as I explain below, I limited my analysis to public-school students who reported only being suspended once or twice rather than multiple times. These students were therefore unlikely to be serial "troublemakers" or to pose a severe threat to staff or other students. Given the frequency with which suspension in used in American public schools (Wallace et al., 2008), I believe

it is appropriate to assume that students receiving one or two suspensions did not commit egregious acts that necessitated school removal.

In addition to assuming that each individual has potential outcomes associated with a treatment that can be manipulated and that his or her probability of receiving the treatment was greater than zero and less than one, the stable unit treatment value assumption (SUTVA) is necessary in order to estimate causal effects in a straightforward manner (Rubin, 2005). First, under SUTVA, the treatment that one individual receives does not affect the potential outcomes of other individuals. Second, one's potential outcomes under treatment and control are the same no matter how treatment is administered. The Education Longitudinal Study of 2002 (ELS:2002) data are characterized by a hierarchical structure: Participating students are nested in their schools. Theoretically, SUTVA may be violated in clustered settings (Hong & Raudenbush, 2003). For example, one can imagine a student who would normally attend college whether or not she was suspended; perhaps, hypothetically, if her best friend were suspended, this might affect her own potential outcome under suspension. Unlike with vaccination and herd immunity, for example, there is no strong theoretical basis that suggests interference occurs to a strong degree.

However, if suspension leads to poorer academic achievement, this could hypothetically affect other students' academic achievement and therefore their potential outcomes. Indeed, there is evidence of negative spillover to non-suspended students academic achievement in a longitudinal study of public middle and high schools in a school district in Kentucky (Perry & Morris, 2014); the authors controlled for non-suspended students' infractions and school-level drug, violence, and disruptive infractions, and placebo tests indicated that this relationship was not driven by unobserved confounding. In the present data set, however, school-level mean math and reading achievement (at baseline) was not statistically significantly related to college matriculation after controlling for students' own academic achievement and other covariates. This suggests that even if suspension affected students' academic outcomes in the present data set, any resulting spillover was not strong enough to affect students' chances of attending college.

Although a SUTVA violation is possible, as is the case with the vast majority of social research, the decision to attend college is multidetermined and related most closely to an individual's academic achievement and socioeconomic status rather than that of his or her peers. Interventions that have dramatically increased college matriculation have focused on the logistics of the college application process rather than peer effects or school context (Bettinger et al., 2012). Therefore, the suspension of other students may not play a substantial role in determining an individual's potential outcomes.

With clustered data, it has been argued that matching is ideally conducted within each cluster (e.g., Rickles & Seltzer, 2014). In practice, however, this may significantly reduce the size of the analyzed sample and exclude "substantively important subgroups" (Rickles & Seltzer, 2014, p. 612). It has also been demonstrated that both within- and across-cluster matching can produce unbiased estimates under certain conditions (Steiner, Kim, & Thoemmes, 2012). Furthermore, it is possible that the closest match (in terms of X_i distance) for a given treated unit is located in another cluster. This presents a trade-off, especially as "schools can and do respond in different ways to different students within

the same schools" (Kupchik, 2010, p. 49). Consider two White students who attend the same school. They both have never been retained, live in a two-parent household, and have average academic achievement. However, one student comes from an extremely wealthy household and the other lives below the poverty line. These students may experience their school in fundamentally different ways and be treated and perceived in very different ways by educators and peers.

When matching, therefore, one should not necessarily privilege finding a within-school match over the best match in terms of background characteristics. Moreover, in this specific case, part of what leads to randomness in suspension is similar students' enrollment in different schools with different discipline policies. Although within-cluster matching is often considered to be ideal, here it may increase the chances of confounding due to some unobserved individual-level characteristic.

On the other hand, schools in the United States vary dramatically with respect to physical conditions, financial resources, and teacher qualifications. Not accounting for these differences could lead to biased estimates if unobserved school characteristics are related both to school suspension rates and college matriculation rates. As I note below, however, my estimates did not change notably whether I matched students within or across schools, and my estimates remained strong and negative after accounting for school characteristics using both random and fixed effects. This pattern of results suggests that the school environment (and therefore the suspension of other students) plays a relatively small role in determining the impact of suspension on educational outcomes.

Taking these considerations into account, I first found the best match based on individual- and school-level characteristics without restricting potential matches to the same school. When matching students across different schools, I took the following school-level variables into account: mean math and reading achievement scores, mean socioeconomic status, geographic region, and urbanicity. I did not match on school suspension rates; differences in administrator beliefs about discipline are a desired source of random variation in similar students being suspended or not. Second, I conducted within-school matching.

4.2 Data Set

The ELS:2002 is an ongoing national longitudinal study of individuals who were enrolled in tenth grade in 2002. During the baseline year, questionnaires were administered to participating students, students' math and English teachers, school principals, and students' parents or guardians. Additionally, students took a standardized academic achievement test in math and reading. Although the ELS:2002 is intended to be a nationally representative study when appropriate weights are used, this analysis focused on a non-representative subset of participants.

This analysis was limited to students who were attending public schools because suspensions are used more rarely in Catholic and other private schools ($n = 3,432$). I believed, *a priori*, that it would be difficult to adequately match suspended students and non-suspended

students in these settings. Even if balance were to be achieved, I believe unobserved confounding poses a greater threat in private schools than in public schools because suspension is more unusual in these settings. I also excluded students who reported being expelled or transferred for disciplinary reasons ($n = 172$) and those who reported receiving more than two in-school suspensions ($n = 426$) or out-of-school suspensions ($n = 207$) during the previous school term. Additionally, there were 1,391 participants who did not report whether they had received an in- or out-of-school suspension.[1] In all, this reduced the number of observations from 16,197 to 10,799 students. [2] Students were nested in and matched according to the school they attended in 2002; participating students attended 579 different schools in the base year. The mean number of students per school was 20 ($SD = 5$), and ranged from 3 to 36 students.

Students who have been suspended many times are more to likely differ from non-suspended students in unobservable ways. Therefore, I aimed to compare students with just one or two instances of harshly punished behavior to students who reported not having been punished in an exclusionary way. Students who reported receiving no in- or out-of-school suspension during the previous term were considered part of the control group. Students who reported receiving one or two in- and/or out-of-school suspensions during the previous term were considered part of the treatment group. This decision to combine in- and out-of-school suspensions was based on a preliminary analysis; the estimated odds ratios associated with both types of suspension were similar with respect to both dropout and college matriculation. By limiting treated students to a) those with only one or two in- and/or out-of-school suspensions and b) those attending public schools, where suspension is commonly used, I minimized the potential influence of unobserved behavioral variables.

Additionally, I controlled for and matched students based on a set of pre-treatment variables that past research has shown are related to suspension or matriculation at a four-year college. These variables included self-identified race/ethnicity (identifying as White served as the reference category); gender (identifying as female served as the reference category); an indicator for participating in sports; an indicator for ever having been retained in a grade; socioeconomic status, which was a composite variable based on parental education, parental occupation, and household income; Individualized Education Program (IEP) status, which signifies that a student was receiving special education services (IEP, no IEP, or IEP status missing, with having an IEP serving as the reference category); an indicator for living in a single-parent household; an indicator for having at least one parent who graduated from college; and a standardized measure of math and reading achievement, which was administered for the purposes of the ELS:2002. Summary statistics for these variables are provided in Tables 6.1 and 6.2. The math and reading achievement scores were centered around the population mean of 50. It should be noted that the standardized measure of math and reading achievement was administered during the spring semester while students were asked to report whether they had been suspended during the "first semester or term" of the school

[1] There were 13 participants who were missing in-school suspension data and 17 students who were missing out-of-school suspension data who were retained for the analysis.

[2] These numbers do not sum exactly because some individuals were part of more than one dropped category, e.g., they were both transferred for disciplinary reasons and reported being suspended more than once or twice.

year. However, this test was intended as an assessment of "cognitive" ability rather than comprehension of material taught during the first semester of tenth grade. Most test questions concerned material that was taught before tenth grade, such as arithmetic and basic operations using decimals and fractions.

After matching, I also checked balance on variables that may be related to educational outcomes but were not initially found to be related to suspension in the data set. These variables included the student's immigration status (first-generation, second-generation, or more than third-generation American); an indicator for whether at least one of the student's grandparents had earned a college degree, which I used as a measure of wealth; an indicator for participating in Advanced Placement classes; an indicator for participating in an International Baccalaureate program; and the number of academic risk factors. Note that the risk factors were whether the student came from a single-parent household, had two parents without a high school diploma, had a sibling who dropped out of school, had changed school two or more times, had repeated at least one grade, or came from a household with an income below the federal poverty line. Some of these variables were included in the matching process directly or indirectly, as part of the socioeconomic variable. I excluded the other risk factor variables from the matching process because existing research does not suggest they are related to suspension and they were reported only by participating parents; many students were missing a parent questionnaire. When checking balance with respect to these other variables, I imputed missing values based on the overall mean or proportion.

At the time of the first follow up (when data on dropping out and other potential mediators were collected), 8% of the base-year participants did not complete a questionnaire. Fourteen percent of students who were suspended at baseline did not respond compared to 7% of students who were not suspended at baseline. At the time of the second follow up (when data on college matriculation were collected), 13% of the base-year participants did not complete an interview. Seventeen percent of students who were suspended at baseline did not respond compared to 11% of students who were not suspended at baseline. These proportions include students who were ultimately dropped from the analysis (e.g., students attending Catholic schools and students who reported being transferred for disciplinary reasons).

In addition to the participants who did not complete any questionnaire or interview at the time of the follow-up, certain participants chose to skip individual questions. Students were excluded from the main analyses if they were missing covariates or the relevant outcome. After dropping students who were missing the suspension outcome, many participants were missing data with respect to two covariates: whether they qualified for an IEP or had ever been retained in a grade. More than 5,000 participants were missing the IEP variable and some 2,000 participants were missing the grade retention variable. The IEP information was obtained from school enrollment lists or school personnel, and the retention information was provided by the parent. Having an IEP was not statistically significantly related to suspension after controlling for other baseline variables in this data set. Forty-five percent of suspended students were missing the IEP status variable compared to 48% of non-suspended students. When IEP status was not included in the matching process, however, unequal numbers of suspended and non-suspended students had an IEP after matching. Therefore, I controlled for and matched students on an indicator for not having an IEP and an indicator

for IEP status not being reported. Retention was statistically significantly related both to suspension and college matriculation after controlling for other variables. Twenty-six percent of suspended students were missing the retention status variable compared to 18% of non-suspended students. Because of the clear positive association between being retained and being suspended and the clear negative association between being retained and matriculating at a four-year college, I chose to retain the retention variable in my analysis despite the amount of missing data. This resulted in a sample of 8,506 students for whom no covariate or suspension data was missing and 7,626 students for whom no covariate, suspension, or college matriculation data was missing.

Removing students who attended private schools, were suspended more frequently, were ever transferred for disciplinary reasons, or were missing data reduces the size of my data set and renders my sample non-representative. However, in eliminating these observations, I eliminate heterogeneity that may bias my estimates (Keele, 2015). A larger sample is not necessarily more desirable in an observational study if the sample's heterogeneity increases along with its size: "increasing the sample size can shrink the confidence intervals to a point that excludes the true treatment effect point estimate" (Keele, 2015, p. 325).

4.3 Data Analysis

4.3.1 Bounds

Treatment effects can be bounded when dealing with binary outcomes (Manski, 1990; Morgan & Winship, 2015). When an outcome is binary, such as attending or not attending a four-year college, an individual's potential outcomes must be either 0 or 1. Therefore, the treatment effect must be between -1 and 1. After observing randomly selected data, one can place further bounds upon the treatment effect. To estimate one bound, one can assume that all of the treated individuals would have had an outcome equal to 1 if they had not been treated, and that all of the control individuals would have had an outcome equal to 0 if they had been treated. To estimate the other bound, one can assume that all of the treated individuals would have had an outcome equal to 0 if they had not been treated, and that all of the control individuals would have had an outcome equal to 1 if they had been treated. The length between the upper and lower bounds will always be equal to 1 and will consequently always include 0. However, this "no-assumptions" bound may still be informative in some sense; it can show how likely it is that the effect is positive or negative.

The bounds on the treatment effect δ are calculated in the following way, where π represents $\mathbb{E}[D]$, or the the probability of being treated (Morgan & Winship, 2015):

$$\begin{aligned}\mathbb{E}[\delta] =&\{\pi\mathbb{E}[Y(1) \mid D=1] + (1-\pi)\mathbb{E}[Y(1) \mid D=0]\} \\ &- \{\pi\mathbb{E}[Y(0) \mid D=1] + (1-\pi)\mathbb{E}[Y(0) \mid D=0]\}\end{aligned} \tag{4.2}$$

This bound can be narrowed if certain assumptions are justified. First, one can narrow the

bound if it can be assumed that the treatment effect is monotonic (Manski, 1997). In this instance, it seems reasonable to assume that there are no students who would attend college if they were suspended but would not attend college if they were not suspended. Although some researchers have argued that suspensions may have a positive effect on peers (e.g. Kinsler, 2013), it is highly unlikely that a student would be more likely to attend college after having been suspended him- or herself, i.e., having missed instruction and having the history of suspension entered into his or her school record. Although Li (2016) proposed that suspension may function as a "wake-up call" that encourages students to behave better in the future, he did not actually find this to be the case in his data set. Given that students who are suspended for offenses receive more office disciplinary referrals in the future than students who are referred but not suspended for offenses (e.g., Tobin, Sugai, & Colvin, 1996), it seems unlikely that suspensions function in this way.[3] Hypothetically, the threat of suspension may dissuade some students from misbehaving in the first place, but there is no empirical evidence, to my knowledge, that demonstrates that being suspended once dissuades students from misbehaving in the future. Second, the bound can be narrowed if a monotonic assumption about selection into treatment can be made (Manski & Pepper, 2000). Here, it is reasonable to assume that those who are not suspended have higher average outcomes than those who are suspended. As reviewed above, there is a large body of literature that demonstrates that suspended students have lower socioeconomic status and lower academic achievement.

4.3.2 Multilevel Modeling

Some statistical techniques such as ordinary least squares regression assume that responses are independent given covariates. This assumption is frequently not met in educational research, where students are nested within the schools they attend (Rabe-Hesketh & Skrondal, 2012). Therefore, multilevel modeling was used to account for the clustering of students within schools in the present study. I used logistic regression models to describe the relationship between my binary outcome variable – matriculating at a four-year college – and student- and school-level explanatory variables.

I fit the multilevel models using Stata 13.1 software and the `xtlogit` command. Stata uses maximum likelihood by adaptive Gaussian quadrature to fit the mixed-effects logistic model. I used 30 integration points to ensure the accuracy of the quadrature approximation. As described above, a student i is nested in school j. I included a random intercept ζ_j to account for dependence within schools. The model can be written in the following way, where $\zeta_j \sim N(0, \psi)$:

$$logit\left\{Pr(y_{ij} = 1 \mid \mathbf{x}_{ij}, \zeta_j)\right\} = \alpha + \beta'\mathbf{x}_{ij} + \zeta_j \tag{4.3}$$

[3]Controlling for the number of referrals and having been referred for harassment during the first term of sixth grade, *not* having been suspended was associated with fewer referrals in subsequent terms (Tobin et al., 1996). In other words, students who were suspended had more referrals in subsequent terms. This study focused on a small number of students, but it strengthened by its longitudinal design.

The school-specific intercepts represent the unobserved school-level covariates that result in some schools producing students who are more likely to enroll in a four-year college. By using mixed-effects logistic regression, school-specific (conditional) odds ratios are estimated instead of population-averaged (marginal) odds ratios.

4.3.3 Matching

Following the recommendations of Morgan and Winship (2015), I estimated the treatment effect with multiple matching techniques. Different matching methods can produce significantly different point estimates and may produce estimates that are significantly above or below the true treatment effect in finite samples (Morgan & Winship, 2015). Therefore, Morgan and Winship (2015) recommend that researchers check whether the estimated treatment effect is similar across multiple matching methods and report these multiple estimates.

4.3.4 Propensity Score Matching

Propensity score matching is based on the idea of modeling an individual's hypothetical likelihood or propensity for being treated using pre-treatment covariates. The goal is then to find pairs of individuals who had a similar propensity for being treated but in actuality consist of one treated individual and one control individual. The propensity score reduces the multiple dimensions associated with many covariates to a single one—the estimated propensity score. Matching on the propensity score should balance the observed covariates that were used to estimate the score itself.

Propensity scores were also estimated using multilevel models in Stata, using the `melogit` command. Predictions were based on the fixed effects estimates and the posterior means of the school-level random effects. The probability of receiving one or two suspensions was estimated using the following individual-level covariates: self-identified race/ethnicity (White served as the reference category), an indicator for sports participation, gender (female served as the reference category), IEP status (having an IEP served as the reference category), standardized math and reading achievement, socioeconomic status, an indicator for having been retained one or more times, an indicator for being eligible for special education services, an indicator for living in a single-parent household, and an indicator for having at least one parent who graduated from college. The following school-level variables were also included: urbanicity (suburban served as the reference category), region (the West served as the reference category), mean standardized math and reading achievement, and mean socioeconomic status.[4] Size of the tenth grade class was included because of the sampling design (Snijders & Bosker, 2012).

As displayed in Figure 6.5, there was sufficient overlap between the estimated propensity

[4]This predicted propensity score was highly correlated with one not including school-level variables, $r = 0.994$.

scores of treated and control students to conduct propensity score matching. For non-suspended students, predicted propensity scores ranged from 0.01 to 0.69. For suspended students, predicted propensity scores ranged from 0.02 to 0.67. As expected, the mean predicted propensity score for suspended students (0.20) exceeded that of non-suspended students (0.11).

First, students were matched based on propensity score distance across the data set; no restriction on within-school matching was imposed. Second, exact matching on the school identifier variable was used to match treated students only to control students who attended the same school. The treatment effects based on propensity score and genetic matching were estimated based on the difference in means, using the Matching package in R (Sekhon, 2011). Abadie-Imbens standard errors were used (Abadie & Imbens, 2006). Non-suspended individuals who were not matched to suspended individuals were pruned from the analysis. Matches were not be pruned based on any caliper (King & Nielsen, n.d.). Matching was conducted with replacement; ties were allowed, and the matched data was weighted accordingly. When testing continuous variables for balance, 1,000 bootstrap samples were used for the Kolmogorov-Smirnov test. The Kolmogorov-Smirnov test non-parametrically tests whether two cumulative distribution functions are the same.

4.3.5 Genetic Matching

In contrast to propensity score matching, which reduces the multidimensional covariate space to a single score, genetic matching retains the original multivariate structure of the data. An evolutionary search algorithm is used to find the covariate weights and matches that most reduce the distance between treated and control units.

Multivariate genetic matching was conducted using the Matching package in R (Sekhon, 2011). Students were matched with respect to the following individual-level variables: self-identified race/ethnicity (Black, Asian, or other), gender, sports participation, standardized math and reading achievement, IEP status (IEP, no IEP, or IEP status unknown), socioeconomic status, having been retained one or more times, living in a single-parent household, and having at least one parent who graduated from college. The choice to match students based on their identification as Black or African American or some other race/ethnicity was grounded in sociological research that suggests this divide is especially important for life outcomes (J. Lee & Bean, 2004). I also included self-identification as Asian because Asian students were statistically significantly less likely to be suspended than students who identified as White after controlling for other variables related to treatment status. Positive stereotypes about Asian Americans are widespread in American society and affect teacher perceptions of students (Diamond, Randolph, & Spillane, 2004). I also matched students based on the following school-level variables: region (Northeast, South, Midwest, or West), urbanicity (suburban, urban, or rural), mean math and reading achievement, and mean socioeconomic status. The propensity score itself was not included as a variable to be matched upon. This decision was based on the possibility of propensity score matching leading to biased estimates when the true data-generating process is not known (King & Nielsen, n.d.).

First, students were matched based on multivariate distance across the data set; no restriction on within-school matching was imposed. Second, exact matching on the school identifier variable was used to match treated students only to control students who attended the same school. As with propensity score matching, matching was conducted with replacement and ties were allowed. The population size, which is the number of individuals used during the optimization process, was set to 1,000. The wait generations was set to 10. When testing continuous variables for balance, 1,000 bootstrap samples were used for the Kolmogorov-Smirnov test.

4.3.6 Coarsened Exact Matching

In coarsened exact matching, matches are conducted exactly. In other words, no imbalance is allowed, and treated units for whom no exact match can be found are discarded. To facilitate this process, continuous, ordinal, and categorical variables may be coarsened based on theoretical reasons. First, exact matching was conducted with respect to students' race (Black, Asian, or neither Black nor Asian), gender, sports participation, retention status, IEP status, living in a single-parent household, parental education, household socioeconomic quartile, and math and reading quartile. As I report below, the estimated odds ratios for the indicators for not having an IEP and not having information about one's IEP status were very similar, both with respect to the treatment and outcome. Therefore, I combined these two groups of students when conducting coarsened exact matching. Second, exact matching was conducted with respect to the aforementioned individual-level characteristics and the following school-level characteristics: school region, school urbanicity, school mean socioeconomic quartile, and school mean math and reading quartile. Exact matching with respect to the aforementioned individual-level characteristics and school attended was not possible due to the high number of dimensions. After exact matching was conducted, the treatment effect was estimated by regressing the outcome on treatment and the variables that were coarsened during the matching process. I used linear regression in order to obtain predicted probabilities that were comparable to the difference in means (or proportions) that were calculated during propensity score and genetic matching. Coarsened exact matching was conducted using the cem package in Stata (Iacus, King, & Porro, 2009).

4.4 Sensitivity Analysis

As described above, the goal of matching is to approximate a randomized experiment. Within matched pairs, the individual who is treated is assumed to be as-if randomly assigned. However, in an observational study, it is possible that matched pairs remain different if an important predictor of treatment goes unobserved. In this case, matched units would have a different propensity for being treated, and the matching will not have approximated a randomized experiment. Unfortunately, this cannot be tested. I matched students based on a set of variables thought to be related both to suspension and college matriculation. I attempted to exclude serial "troublemakers" by limiting my analysis to public school stu-

dents who reported no more than two of either kind of suspension. However, it is possible that matched and non-matched students still differed with respect to self-control, parental involvement, or some other unobserved variable. These unobserved confounders will most pose a problem if they are unrelated to the variables that are observed and used when matching; self-control is presumably related to baseline academic achievement, for example, and some research suggests that "problem" behavior is related to a student's socioeconomic status. Risk of suspension may also be related to individual characteristics such as obedience to authority and passivity, but these characteristics are less likely to be related to college matriculation, and are likely related with observed characteristics such as participation in sports.

I therefore conducted a sensitivity analysis to determine how large a confounder would need to be present to account for the estimated treatment effect (Rosenbaum, 2002, 2010). The statistic used to quantify the potential effect of hidden bias is the ratio of similar (in terms of X_i) treated and control units' odds of being treated. The closer the statistic Γ is to 1, the more vulnerable the findings are to hidden bias (Rosenbaum, 2002, 2010). I tested the sensitivity of the propensity score and genetic matching results using the rbounds package in R (Keele, 2010).

Chapter 5

Results

5.1 Descriptive Findings

Schools varied widely in how frequently they suspended students, as depicted in Figure 6.2. The proportion of participating students who were suspended ranged from 0 to 60% across schools. The mean suspension rate was 13% (SD = 11%). The median suspension rate was 12%, the 25th percentile was 6%, and the 75th percentile was 20%. Likewise, schools varied widely according to what proportion of students eventually attended a four-year college, as depicted in Figure 6.3. The proportion of students attending a four-year college ranged from 0 to 100% of sampled students across schools. The mean college attendance rate was 44% (SD = 22%). The median college attendance rate was 42%, the 25th percentile was 29%, and the 75th percentile was 57%. The significant variation in school suspension and college matriculation rates suggested that it was important to consider the impact school characteristics might have had on any apparent association between suspension and college matriculation. This was accomplished through multilevel modeling and by matching students both within and across schools.

At the school level, school suspension rate and college matriculation rate were negatively correlated, $r = -0.303$, $p < 0.001$. However, this relationship was not necessarily linear. In fact, as depicted in Figure 6.4, the relationship between suspension and college matriculation seemingly became positive after approximately 40% of sample students were suspended. Only a small number of schools suspended students at such high rates ($n = 21$). Therefore, a continuing negative relationship could not be ruled out, as is depicted in the gray 95% confidence interval surrounding the local regression plot.

The following characteristics were associated with greater risk of suspension, after controlling for other covariates: Black/African American ethnicity, identifying with more than one race/ethnicity (excluding Hispanic/Latino), male gender, history of retention, and living in a single-parent household. Asian American or Pacific Islander ethnicity, participation in sports, and standardized academic achievement were negatively associated with suspension risk. Household socioeconomic status and having a parent who had graduated from college were

negatively associated with suspension risk, although the two variables were not statistically significant when both were included in the model.[1] These results are summarized in Table 6.5. Twenty-six percent of Black students reported being suspended compared to only 11% of White students, $\chi^2(1, N = 7,322) = 235.29, p < 0.001$. Forty-seven percent of students who were not suspended matriculated at a four-year college compared to 20% of suspended students, $\chi^2(1, N = 9,432) = 329.335, p < 0.001$. Black students also remained more likely to be suspended after including fixed effects for schools.

5.1.1 Bounds

The following bounds calculations are based on the expectation that 14% of the population is suspended, or treated.

Under no assumptions about the direction of treatment selection or treatment response, the upper bound for the average treatment effect on the treated (ATT) is 0.20, and the lower bound is -0.80. The upper bound for the average treatment effect (ATE) is 0.48, and the lower bound is -0.52. These results are summarized in Table 6.3.

Based on the research reviewed above, it is reasonable to assume that students who are not suspended have higher average potential outcomes under treatment and control than those who are suspended: $E[Y(1) \mid D = 0] \geq E[Y(1) \mid D = 1]$ and $E[Y(1) \mid D = 1] \geq E[Y(1) \mid D = 0]$. Under this monotonic treatment selection assumption, the lower bound shifts to -0.27. Additionally, the vast majority of educational research supports the hypothesis that the individual-level treatment effect is not positive: $\delta \leq 0$ for every individual i. Under this monotonic treatment response assumption, the upper bound shifts to 0. These results are summarized in Table 6.4.

5.2 Multilevel Modeling

The likelihood-ratio tests for the null hypothesis that the residual between-cluster variance ψ is zero indicated that multilevel models were appropriate ($p < 0.001$). The intraclass correlation for latent responses, which represents the proportion of total variance that is shared amongst students in the same schools, was estimated to be 0.05 for suspension and 0.08 for matriculation at a four-year college. This degree of correlation is typical for educational studies (Hedges & Hedberg, 2007).

After controlling for race, gender, participation in sports, math and reading achievement at baseline, retention status, IEP status, living in a single-parent household, household socioeconomic status, parental education, and school characteristics, suspension remained strongly negatively associated with matriculating at a four-year college, $OR = 0.454$, 95% CI [0.359,

[1] I retained both in the model used to estimate the propensity score and as matched characteristics because both variables were statistically significant with respect to the college matriculation outcome.

0.576], $p < 0.001$.[2] After controlling for suspension and other covariates, Black students were more likely to matriculate at a four-year college than White students, $OR = 2.429$, 95% CI [1.961, 3.008], $p < 0.001$. Additionally, I estimated the association between suspension and college matriculation for Black students alone to examine whether effect heterogeneity was present. The estimated relationship between suspension and college matriculation for this subset of students was similar to that of the entire sample, $OR = 0.508$, 95% CI [0.331, 0.781], $p = 0.002$. As explained above, these estimates are relevant for students who attend the same school or schools with identical random effects. All estimated odds ratios and their 95% confidence intervals after controlling for student and school characteristics are displayed in Table 6.6.

In order to check the robustness of these estimates, I also fit a fixed effects logistic regression to fully control for the potential influence of students' schools. Random effects estimates will not be consistent if the unobserved, school-level heterogeneity ζ_i is correlated with observed covariates $\mathbf{x}_{ij}$ (Gardiner, Luo, & Roman, 2009). The estimated association between suspension and college matriculation remained strongly negative and statistically significant in the fixed effects logistic regression, after controlling for other covariates, $OR = 0.448$, 95% CI [0.361, 0.555], $p < 0.001$. The strength and direction of other estimates did not change markedly. The results of the fixed effects logistic regression are also summarized in Table 6.6.

As described above, the math and reading achievement test was administered in the spring term, and students were asked whether they had been suspended during the fall term. Although the assessment was not intended to focus on material taught during the preceding term, it is technically a post-treatment variable. Therefore, I fit the mixed-effects multilevel model a second time without the participant- and school-level achievement variables. Without these two variables, the estimated effect of being suspended on college matriculation remained strongly negative and statistically significant, $OR = 0.344$, 95% CI [0.276, 0.429], $p < 0.001$. Therefore, it does not appear that the negative effect of suspension is driven by inclusion of a post-treatment variable. Given the similarity of these estimated effects and the greater risk of confounding due to omitted variable bias, all of the matching analyses were conducted with the math and reading achievement variable.

I also examined whether these results were sensitive to the choice of matriculation at a four-year college, rather than a two-year college or greater, as the outcome. As I discuss above, approximately one third of students who initially enroll at a two-year college transfer to a four-year college or university (Jenkins & Fink, 2016). The results of the above mixed-effects model did not change substantially with respect to my main variables of interest when I included individuals who matriculated at a two-year college ($n = 2,463$) along with those who matriculated at a four-year college. Suspension remained strongly negatively associated with attending a college, $OR = 0.532$, 95% CI [0.445, 0.636], $p < 0.001$. Additionally, students who identified as Black or African American remained more likely to attend a college than students who identified as White, after controlling for other covariates, $OR = 1.831$,

[2] As expected, the strength of the association between suspension and matriculation at a four-year college was attenuated after including student- and school-level covariates, but it remained strongly negative. The estimated odds ratio for suspension in the multilevel model with no controls was 0.264, 95% CI [0.221, 0.315].

95% CI [1.471, 2.279], $p < 0.001$. When I limited my analysis to students who did not attend a four-year college, suspension was strongly negatively associated with attending a two-year college, controlling for other covariates, $OR = 0.663$, 95% CI [0.558, 0.789], $p < 0.001$.

Finally, I re-fit the mixed-effects model using the participants who were matched using genetic matching across schools, which I describe below. The estimated association between suspension and matriculation at a four-year college remained strong, negative, and statistically significant in this reduced sample, after controlling for other covariates, $OR = 0.445$, 95% CI [0.335, 0.590], $p < 0.001$.

5.3 Propensity Score Matching

When matching on the propensity score across the data set, all 886 suspended students were matched with 4,732 different non-suspended students. The maximum number of times a control unit was used as a match was eight times. After matching, t-tests indicated that the mean of the standardized math and reading achievement score (-4.63 vs. -5.55) was statistically significantly different between the treatment and control groups. Additionally, Kolmogorov-Smirnov tests indicated that the distributions of math and reading achievement ($p < 0.001$) and socioeconomic status ($p < 0.001$), as well as the distributions of these school-level variables, remained statistically significantly different. The degree of balance before and after matching is summarized in Table 6.7 and Figure 6.6. As explained above, I also checked balance on variables that were not used to estimate the propensity score; missing values were imputed based on the overall mean or proportion. The degree of balance before and after matching is summarized in Table 6.8. After matching, t-tests indicated that only variable that was statistically significantly different between the treated and control students was related to the treatment itself. Suspended students attended schools with higher suspension rates (22%) than non-suspended students (16%). The estimated SATT of suspension on matriculating at a four-year college was -0.078 ($AISE = 0.017$).

When matching treated students only to other students from their school, all 886 suspended students were matched to 752 different non-suspended students. The maximum number of times a control unit was used as a match was four times. After matching, t-tests indicated that the following covariates used to estimate the propensity score were statistically significantly different between the treatment and control groups: math and reading achievement (-4.63 vs. -3.97), proportion retained (24% vs. 16%), proportion with an IEP (12% vs. 9%), and proportion with a college-graduate parent (26% vs. 22%). The degree of balance before and after matching is summarized in Table 6.9 and Figure 6.8. Additionally, t-tests indicated that several variables not used to estimate the propensity score were imbalanced after imputing missing values; these results are summarized in Table 6.10. More suspended students (85%) spoke English as their native language than non-suspended students (78%). Fewer suspended students (8%) than non-suspended students (10%) were first-generation immigrants. Fewer suspended students (10%) than non-suspended students (13%) were second-generation immigrants. Suspended students had slightly more risk factors (1.5) than non-suspended students (1.3). More suspended students (85%) had parents who spoke En-

glish as a native language than non-suspended students (81%). The estimated effect of suspension for the treated students who were matched was -0.124 ($AISE = 0.021$), but this estimate likely reflects the remaining imbalance between the two groups. However, some of this bias was downwards. For example, after matching, more suspended students than non-suspended students had at least one parent who was a college graduate.

5.4 Genetic Matching

When matching on all covariates except school, all 886 suspended students were matched with 757 different non-suspended students. The maximum number of times an observation was used as a match was six times. After matching, t-tests indicated that none of the means of the covariates used during the matching process were statistically significantly different between the treatment and control groups. However, more suspended students (19%) identified as Hispanic or Latino than non-suspended students (16%). As described above, individuals were matched according to whether they identified as Asian, Black, or other rather than each individual race/ethnicity. The degree of balance before and after matching is summarized in Table 6.11 and Figure 6.10. Additionally, t-tests indicated that three variables not used in the matching process were statistically significantly different between the suspended and non-suspended students, as summarized in Table 6.12. Suspended students had slightly more risk factors (1.5), on average, than non-suspended students (1.4). Suspended students attended schools with slightly lower college matriculation rates (38%) than non-suspended students (40%). Suspended students attended schools with higher suspension rates (21%) than non-suspended students (16%). The estimated SATT of suspension on matriculating at a four-year college was -0.113 ($AISE = 0.021$).

When matching students with other students from their school, all 886 suspended students were matched to 745 different non-suspended students. The maximum number of times an observation was used as a match was five times. After matching, t-tests indicated that the following were statistically significantly different between the treatment and control groups: proportion of Asian students (4% vs. 3%), the proportion of Black students (25% vs. 22%), the proportion of male students (55% vs. 49%), the proportion of students who participated in sports (44% vs. 49%), mean math and reading achievement (-4.63 vs. -3.91), the proportion of retained students (24% vs. 17%), the proportion of students living in a single-parent household (34% vs. 25%), the proportion of students who did not have an IEP (44% vs. 46%), the proportion of students who were missing information about their IEP status (44% vs. 47%), and the proportion of students who had at least one parent who graduated from college (26% vs. 22%). Additionally, the Kolmogorov-Smirnov test indicated that the two groups' distributions of math and reading achievement ($p = 0.042$) remained statistically significantly different. The degree of balance before and after matching is summarized in Table 6.13 and Figure 6.12. Additionally, t-tests indicated that several other variables not used for matching were imbalanced, as summarized in Table 6.14 and Figure 6.13. More suspended students were native English speakers (85%) than non-suspended students (82%). Fewer suspended students (10%) than non-suspended students (12%) were second-generation immigrants. Fewer suspended students (19%) than non-suspended students (23%) had at least

one grandparent who graduated from college, a proxy for family wealth. Fewer suspended students were enrolled in Advanced Placement classes (12%) than non-suspended students (15%). Suspended students had more risk factors (1.5), on average, than non-suspended students (1.2). The estimated effect of suspension for the treated students who were matched was -0.144 ($AISE = 0.021$). However, this estimate likely reflects bias due to the remaining imbalance between the suspended and non-suspended students.

5.5 Coarsened Exact Matching

Because not all treated units were matched through the coarsened exact matching, the following estimates represent the local SATT (Iacus, King, & Porro, 2012). Descriptive statistics of the matched samples are provided in Tables 6.15 and 6.16.

After conducting exact matching with respect to all individual-level covariates, 806 out of 886 suspended students (91%) and 4,816 out of 6,740 non-suspended students (71%) were matched. Out of the 1,001 strata containing observations, 355 strata were matched.[3] The estimated effect of suspension on matriculation at a four-year college was -0.102, 95% CI [-0.133, -0.072], $p < 0.001$. African American students were approximately 17% more likely to attend college than White students, 95% CI [0.146, 0.203], $p < 0.001$, controlling for other covariates.

After conducting exact matching with respect to all individual- and school-level covariates (but not school itself), 214 out of 886 suspended students (24%) and 382 out of 6,740 non-suspended students (6%) were matched. Out of the 5,791 strata containing observations, 199 were matched. The estimated effect of suspension on matriculation at a four-year college was -0.100, 95% CI [-0.167, -0.034], $p = 0.003$. African American students were approximately 18% more likely to attend college than White students, 95% CI [0.074, 0.296], $p = 0.002$, controlling for other covariates.

5.6 Sensitivity Analysis

The results of this sensitivity analysis varied somewhat according to the type of matching conducted. For propensity score matching across schools, the p-value would no longer be less than 0.05 for $\Gamma = 1.8$, i.e., if suspended students were 1.8 times more likely to be suspended than their non-suspended matches due to some unobserved confounder. For propensity score matching within schools, the p-value would no longer be less than 0.05 for $\Gamma = 1.8$. For genetic matching across schools, the p-value would no longer be less than 0.05 for $\Gamma = 1.6$. For genetic matching within schools, the p-value would no longer be less than 0.05 for $\Gamma = 1.9$.

[3] An example of un-matched stratum was a male Asian and/or Pacific Islander student who played a sport, was in the second quartile for math and reading achievement, was in the third quartile for household socioeconomic status, was never retained, lived in a household headed by a single parent, had a parent who graduated from college, was not reported as having an IEP, and who was suspended.

5.7 Exploratory Analysis of Causal Mechanisms

I also conducted an exploratory analysis of potential causal mechanisms. In order to explore what factors explain suspended students' lower chances of attending a four-year college, I first examined the bivariate associations between suspension and potential mediating variables. These apparent associations could be due to confounding (e.g., the association between lower socioeconomic status and suspension) rather than a true causal relationship. Therefore, I then examined the relationship between each potential mediator and the treatment after controlling for the baseline variables used in the main analyses described above. I then examined the relationship between each potential mediator and the treatment after controlling for the baseline variables used in the main analyses described above. Mixed effects multilevel models were used to account for the clustering of students within schools. If not all important baseline covariates were observed, these associations may reflect unobserved confounding instead of or in addition to a mediational association.

5.7.1 High School Dropout

At the school level, suspension rate and dropout rate were weakly positively correlated, $r = 0.103$, $p = 0.013$.

Students who were suspended at baseline were more likely to drop out from school. Six percent of students who were not suspended dropped out from high school at some point compared to 19% of suspended students.

Suspended students were also more likely to explicitly report that they dropped out due to being suspended at some point. In the follow-up survey of participants who had dropped out, 18% of participants who were suspended during the first term of tenth grade reported eventually leaving school because they were suspended compared to 8% of students who were not suspended. This difference was statistically significant, $\chi^2(1, N = 455) = 8.342, p = 0.004$. Students who were suspended at baseline were also more than twice as likely – 11% versus 5% – to report having dropped out because they were eventually expelled, $\chi^2(1, N = 454) = 5.619, p = 0.018$.

After controlling for baseline student- and school-level covariates, suspended students were much more likely to have ever dropped out from school, $OR = 2.991$, 95% CI [2.362, 3.786], $p < 0.001$. The association between suspension and matriculation at a four-year college remained strong and negative after controlling for dropping out and baseline covariates, $OR = 0.512$, 95% CI [0.402, 0.652], $p < 0.001$.

5.7.2 Reduced Academic Achievement

At the time of the first follow up, students who were suspended had math[4] scores that were seven points lower ($M = 43.9, SD = 8.6$) than non-suspended students ($M = 50.6, SD =$

[4]Reading achievement was not tested at the time of the first follow up.

9.9). This difference was statistically significant, $t(1332.87) = 22.893, p < 0.001$.

After controlling for baseline math and reading achievement and other demographic variables, students who were suspended at baseline had lower math achievement at the time of the first follow up, $\hat{\beta} = -0.951$, 95% CI [-1.317, -0.586], $p < 0.001$.[5] This effect is relatively small, however, as it is equal to approximately one tenth of a standard deviation on the test. The association between suspension and matriculation at a four-year college remained strong and negative after controlling for follow-up math achievement and baseline covariates, $OR = 0.582$, 95% CI [0.449, 0.755], $p < 0.001$.

5.7.3 Peer Influence

At the time of the second follow-up, students reported whether none, a few, some, most, or all of their friends planned to attend a four-year college or university; students also had the option of reporting that they did not know. I created an indicator that was equal to 1 if a student reported that "most" or "all" of their friends planned to attend college and 0 otherwise. Fifty-two percent of non-suspended students reported that most or all of their friends planned to attend college compared to only 32% of suspended students, $\chi^2(1, N = 9,784) = 175.419, p < 0.001$. After controlling for baseline student- and school-level covariates, suspended students were much less likely to report that most of their friends planned to attend a four-year college or university, $OR = 0.653$, 95% CI [0.551, 0.774], $p < 0.001$. After controlling for baseline variables and this peer influence indicator, the relationship between suspension and college attendance remained strongly negative and statistically significant, $OR = 0.507$, 95% CI [0.396, 0.648], $p = 0.001$.

5.7.4 College Admissions

I examined whether suspension at baseline was associated with taking college admissions tests, choosing to apply to college, or the number of colleges that accepted a student. If suspended students are viewed as "troublemakers" who are not "college material," teachers and school counselors may be less likely to encourage these students to take college admissions tests or to apply to college. Furthermore, many colleges ask in their applications whether students have been suspended. Knowledge of this may dissuade students with a history of suspension from taking the steps to apply to college. A suspension may also act as a red flag to college admissions boards, similar to a criminal conviction on a job application (Pager, Western, & Bonikowski, 2009).

Only 35% of suspended students reported taking college admissions tests compared to 66% of non-suspended students. This difference was statistically significant, $\chi^2(1, N = 10,799) = 518.65, p < 0.001$. After controlling for baseline student- and school-level covariates, suspended students were much more less likely to have taken college admissions tests,

[5] The size and statistical significance of this coefficient are similar when baseline math achievement rather than math and reading achievement combined was controlled for.

$OR = 0.404$, 95% CI [0.336, 0.486], $p < 0.001$. The association between suspension and matriculation at a four-year college remained strong and negative after controlling for taking admissions tests and baseline covariates, $OR = 0.646$, 95% CI [0.499, 0.838], $p = 0.001$.

Only 61% of suspended students ever applied to college or another form of postsecondary school compared to 82% of non-suspended students.[6] This difference was statistically significant, $\chi^2(1, N = 9,373) = 302.141, p < 0.001$. After controlling for baseline student- and school-level covariates, suspended students were much less likely to have ever applied to college, $OR = 0.453$, 95% CI [0.379, 0.542], $p < 0.001$. The association between suspension and matriculation at a four-year college remained strong and negative after controlling for ever having applied to college and baseline covariates, $OR = 0.583$, 95% CI [0.453, 0.749], $p < 0.001$.

Amongst participants who applied to college, students who were suspended at baseline were accepted on average by 1.27 colleges whereas students who were not suspended were accepted on average by 1.86 colleges.[7] Twenty-three percent of suspended students were not accepted to any of the colleges they applied to compared to 11% of students who were not suspended. The 25th percentile and median value was one college acceptance for both groups, whereas the 75th percentile was three college acceptances for the non-suspended participants and two college acceptances for the suspended participants. After controlling for baseline student- and school-level covariates, suspended students were more likely to have been accepted to no colleges, but this difference was not statistically significant at the pre-specified alpha level, $OR = 1.379$, 95% CI [0.999, 1.904], $p = 0.051$. The association between suspension and matriculation at a four-year college was negative but no longer statistically significant after controlling for have been accepted to no colleges and baseline covariates, $OR = 0.809$, 95% CI [0.587, 1.116], $p = 0.196$.[8]

[6]The inconsistency between the application rate and the college admissions test taking rate may be due to these two items being collected at different times. Participants reported whether they had taken a college admissions test at the time of the first follow up, and they reported whether they had ever applied to college at the time of the second follow up.

[7]Seventeen students reported that they were accepted by "9 or more" colleges. This value was treated as equal to 9 for the purpose of calculating the mean.

[8]Although not being accepted to college should very closely predict who matriculates at a four-year college and could potentially explain the influence all other variables, the relationship between matriculation and many other baseline covariates (e.g., history of retention, math and reading achievement, and socioeconomic status) remained statistically significant in this model. Some students were accepted by a college but did not matriculate. Some students were classified as not being accepted to a college even though they reported attending a four-year college because colleges with "open-enrollment" policies were excluded from this acceptance item.

Chapter 6

Discussion

6.1 Summary and Analysis

In this dissertation, I examined the association between high school suspension and college matriculation with the aim of determining whether the disproportionate suspension of African American students contributes to the racial gap in educational outcomes. With few exceptions, past research on the association between exclusionary discipline and educational outcomes has failed to account for baseline differences in academic achievement. Additionally, almost all research that has accounted for the potential influence of socioeconomic status has done so through students' free or reduced-price lunch status; this coarse measure does not adequately capture the potential confounding influence of parental education. I controlled for the influence of the following individual-level characteristics to minimize potential confounding in my analysis: self-identified race/ethnicity; gender; having been retained in a grade; socioeconomic status, which was a composite variable based on parental education, parental occupation, and household income; receiving special education services; living in a single-parent household; having at least one parent who graduated from college; a standardized measure of math and reading achievement; and participating in sports. I balanced the trade-off between accounting for unobserved school characteristics and minimizing potential student-level confounding (e.g., unobserved levels of self-control not accounted for by the related variables I did observe) through mixed-effects multilevel models with covariates for influential school-level characteristics, fixed-effects logistic regression to account for observed and unobserved school-level characteristics, and matching students both across similar schools and within the same school. I included the following school-level characteristics in my multi-level models and across-school matching analyses: mean socioeconomic status, mean math and reading achievement, geographical region, and urbanicity.

Suspended students tended to have lower baseline academic achievement than non-suspended students, confirming the suspicion that much research on this topic has overestimated the effect of suspension by failing to account for these differences (Noltemeyer et al., 2015). Even after accounting for these baseline differences, however, suspension appears to have had a strong negative impact on students' educational outcomes. Students who were suspended

were approximately 10% less likely to matriculate at a four-year college across various matching methods, whether or not students were matched within the school they attend. These results are summarized in Table 6.17. Although balance was made worse by matching solely within schools, as is typical (Rickles & Seltzer, 2014), the relative consistency of the estimates across different matching methods suggests that the estimated effect is not an artifact of school context. This effect is present even after controlling for an objective measure of baseline achievement in math and reading, along with potential confounds such as gender, race, socioeconomic status, and history of retention. These are large effect sizes for educational outcomes (Hattie, 2009), although similar-sized effects on college attendance have been documented before (Bettinger et al., 2012).

Suspension was associated with multiple negative intermediate outcomes, after controlling for baseline characteristics. These intermediate outcomes included dropping out from high school, associating with peers who did not plan to attend college, not taking college admissions tests, not applying to college, and being accepted to fewer colleges. I examined these intermediate outcomes in light of existing correlational research on the consequences of school discipline (e.g., Fabelo et al., 2011), long-standing beliefs about the stigmatizing effects of suspension (Children's Defense Fund, 1975), and the necessary steps to gain admission to a four-year college (Perna, 2006; Weissman & NaPier, 2015). In this analysis, I cannot determine whether these associations are due to unobserved confounding, a true mediational association, or both. Investigating these relationships is a direction for future research.

Overall, these findings strongly suggest that increasing suspensions would worsen racial gaps in achievement and attainment rather than ameliorating them, and they refute the conclusions of Kinsler (2013). Given the size of estimated SATT, the estimated effect may reflect future suspensions (i.e., students suspended during this term may have been more likely to be suspended in future terms); being disciplined at one point in time predicts being disciplined in the future (Bowman-Perrott et al., 2013; Theriot, Craun, & Dupper, 2010). An initial suspension may initiate a vicious cycle in which students fall behind, become frustrated in class, are seen as less academically competent or promising by their teachers, and are more likely to truly act out. Additionally, suspensions may aggravate tense relationships between students and their referring teachers, whom students may perceive as unfair or unconcerned with their well-being (Gregory & Weinstein, 2008). Any future suspensions would be considered part of the treatment effect, and the existence of such cycles should be investigated in future research.

6.2 Limitations

This analysis focused on tenth grade students attending public high schools in the United States. These results may not generalize to students in other grades or other countries (Cobb-Clark, Kassenboehmer, Le, McVicar, & Zhang, 2015). Cobb-Clark et al. (2015) studied the effect of receiving a suspension at any point during one's educational career in Australia. Although they found that suspension remained negatively associated high school graduation

or university entrance exam scores after controlling for background characteristics, they concluded that suspension did not negatively impact educational outcomes because of the results of their sensitivity analysis. However, suspension is much more unusual in Australia than the United States, with only approximately 2% of Australian students being suspended in a given year. It should also be noted that the research conducted by Cobb-Clark et al. (2015) was retrospective and controlled for grade retention at any point, a variable that may be post-treatment in their analysis. Additionally, they only analyzed the effect of suspension on university exam test scores for students who completed high school.

The present analysis was based on students' self-reported suspension history. Some students may have intentionally reported inaccurate information about their suspension history, either claiming that they had been suspended when they had not or claiming that they had not been suspended when they had. Additionally, students may have reported what they thought to be true but that did not correspond to other students' interpretations. Suspensions are defined differently at different schools. For example, being sent to a separate classroom for one or two class periods may constitute an in-school suspension in some schools but not others. An objective description of what constituted a suspension was not provided to students on the ELS:2002 questionnaire.

A non-negligible portion of the sample was lost to attrition by the time of the first and second follow-up studies, when outcome data was collected. This attrition could have affected the size of these estimates. Additionally, some statistically and educationally significant differences between suspended and non-suspended students remained after matching students, especially when matching within schools. In a follow-up analysis, I conducted exact matching across schools with respect to the student and parent speaking English as a native language, immigration status, third-generation-college-student status (i.e., the indicator for wealth), enrollment in an Advanced Placement class, and enrollment in an International Baccalaureate program, in addition to the original set of variables, and obtained an estimated effect of suspension of -0.103. Therefore, it does not appear that imbalance is strongly affecting the estimated treatment effect.

As described above, the causal interpretation of these results relies on the assumption that I have observed and condition on the variables that are related both to the treatment and the outcome. Indeed, I controlled for a host of variables that are thought to predict whether students are suspended and enroll in college. For example, unlike many other studies of the effect of suspension, I was able to control for baseline academic achievement and a fine-grained measure of socioeconomic status. However, I did not control for students' self-control, students' educational aspirations, or parent involvement. These variables are likely related to ones that I do observe, but I cannot rule out the possibility that the apparent association between suspension and college matriculation is being driven by some unobserved variable. Future research on the effects of suspension on academic outcomes should take advantage of random variation in suspension induced by exogenous changes in school disciplinary policy to more robustly identify causal effects.

Finally, as these results are based on data collected in 2002, 2004, and 2006, they may not generalize to students enrolled in high school today. Presumably, the effect of suspension could now be larger as the Common Application now asks students to report whether they

have been suspended.

6.3 Implications for Policy and Practice

Given these results, educators and policymakers should ask themselves "Is suspension ever justified?" It may be if it benefits (a) the particular student who is suspended or (b) the student's classmates. Such benefits may arise if suspensions deter future misbehavior or if removing the student results in a classroom environment that is more conducive to learning for the student's peers. However, there is little evidence that suspensions deter future misbehavior since being suspended at one point in time is a good predictor for being suspended in the future. As described above, suspensions most frequently are the consequence of purportedly disrespectful or defiant behaviors (e.g., refusing to remove one's hat (Dominus, 2016)) that most adolescents engage in at least occasionally. Moreover, due to their immature prefrontal cortex, adolescents are unable to weigh potential risks and consequences like adults do (Dahl, 2004; Blakemore & Choudhury, 2006; Steinberg, 2008). Adolescents' inability to accurately assess the risks of their behavior and inhibit risk-taking becomes especially pronounced in the presence of peers (Gardner & Steinberg, 2005; Chein, Albert, O'Brien, Uckert, & Steinberg, 2011), as is the case in school settings. Therefore, few students are likely dissuaded from misbehaving because of the possibility of suspension.

Moreover, students are frequently suspended for mildly defiant behaviors that do not significantly impact their learning or other students' learning or safety. Even if a student has been removed for truly being disruptive or posing a danger to others, a suspension is temporary by nature. Suspensions do not address or remedy underlying causes of misbehavior (Kupchik, 2010); they simply punish the student by removing him or her from the instructional setting or school. Even if there are temporary benefits when a disruptive student is absent, suspension may have a net negative impact on peers if it leads to worse academic outcomes for or behavior from the suspended student in the future.

It appears that costs of suspension may far outweigh any potential benefits, although this study cannot rule out the influence of an unobserved confounder such as self-control. Although unobserved factors like self-control or a "troublemaker" nature may partially be at play, there are clear theoretical reasons to believe that less instructional time and the a disciplinary record would affect students' chances of attending college. Removing students from school as a punishment appears antithetical to schools' primary mission – to educate children.

The negative effect of suspension on individual, family, and community outcomes may be further compounded by other associated outcomes not studied here. When young people drop out from high school or fail to attend college, this impacts their families and communities, perpetuating vicious cycles of disadvantage. For example, high school suspension is associated with involvement in the juvenile justice system (Fabelo et al., 2011). Additionally, individuals who obtain fewer years of education are more likely to become involved in the criminal justice system as adults. A criminal background is then a legal basis for discrimination in employment and housing and for legal disenfranchisement. Having an incarcerated

parent is associated with worse socio-emotional outcomes, poorer academic achievement, and less trust in public institutions among children.

Young people who drop out from high school or who do not attend college are less likely to be employed. Historically, a school record of suspensions could directly impact an adolescents' ability to get a job after dropping out.[1] In turn, unemployed people receive social assistance from the government; such support is necessary but does divert funds. Individuals with low levels of education who do manage to find employment are more likely to work in low-paying jobs, which means that they pay less or no income tax to the government to fund federal programs. Individuals who are less educated are less able to contribute to the modern, knowledge-based economy, where economic growth and the development of beneficial innovations are dependent upon the education of the workforce (Strulik, Prettner, & Prskawetz, 2013). Young people who attend college are more likely to innovate and develop new technologies that will allow people to live longer, healthier lives in a more sustainable world.

Based on the estimated increase in high school drop out alone and presumed causality, suspensions cost the American government and taxpayers billions of dollars each year (Rumberger & Losen, 2016). Individuals with less education tend to have worse health outcomes, which, in addition to impacting quality of life, places financial burdens upon individuals and taxpayers. Although people who obtain a college education differ any many ways from those who do not, the positive association between education and better life outcomes is thought to be causal (Card, 2001; Hout, 2012).

These consequences are especially concerning given that Black students are much more likely to be suspended than White students. These results suggest that discipline is an important factor to consider in the achievement gap, along with differential teacher expectations (Lewis & Diamond, 2015; Weinstein, 2002), persistent school and neighborhood segregation (Card & Rothstein, 2007; Reardon & Owens, 2014), and racialized academic tracking within seemingly integrated schools (Lewis & Diamond, 2015). Suspension may therefore constitute a form of institutional discrimination: "decisions and processes that may not themselves have any explicit racial content but that have the consequence of producing or reinforcing racial disadvantage" (Pager & Shepherd, 2008, p. 182). Like a criminal record (Alexander, 2010), a history of suspension becomes a legal basis for racial discrimination. Disproportionate discipline is therefore one of the ways in which schools perpetuate existing racial inequalities in a supposedly postracial society and contribute to intergenerational cycles of racial inequality in this country (cf. Bourdieu & Passeron, 1964).

6.4 Conclusions

Although this study cannot definitively determine that the negative correlation between suspension and academic outcomes is causal, it provides stronger evidence that suspension

[1]The Children's Defense Fund (1975) documented a student being denied four jobs because of the school suspensions on his record (p. 48). I am unaware of any recent research that addresses this question.

negatively affects students' academic achievement and educational attainment. Even if the relationship between suspension and college matriculation remains confounded by some unobserved variable, these findings clearly demonstrate that suspended students are less likely to attend college. Schools and non-school educational programs that are committed to increasing students' chances of attending college should see suspension as a warning sign and give suspended students additional support.

Long-term suspension or expulsion from school may be the most appropriate response to students who pose an ongoing risk to other students or teachers. In other instances, less harmful consequences should be considered. For example, school-wide positive behavior support systems may both improve academic outcomes and reduce the number of office disciplinary referrals and suspensions (Elfner Childs, Kincaid, Peshak George, & Gage, 2016; R. H. Horner et al., 2009). When a punitive consequence is deemed necessary, a punishment such as after-school detention may be appropriate. Such policy changes have the potential to benefit not only students, who could have greater chances of educational success, but also teachers, who join their profession with the goal of making a positive impact on children's lives.

References

Abadie, A., & Imbens, G. W. (2006). Large sample properties of matching estimators. *Econometrica*, *74*(1), 235–267. doi: 10.1111/j.1468-0262.2006.00655.x

Alexander, M. (2010). *The new Jim Crow*. New York, NY: The New Press.

Alon, S., & Tienda, M. (2005). Assessing the "mismatch" hypothesis: Differences in college graduation rates by institutional selectivity. *Sociology of Education*, *78*, 294–315.

American Psychological Association Zero Tolerance Task Force. (2008). Are zero tolerance policies effective in the schools? An evidentiary review and recommendations. *American Psychologist*, *63*(9), 852–862. doi: 10.1037/0003-066X.63.9.852

Arcia, E. (2006). Achievement and enrollment status of suspended students: Outcomes in a large, multicultural school district. *Education and Urban Society*, *38*(3), 359–369. doi: 10.1177/0013124506286947

Balfanz, R., Byrnes, V., & Fox, J. (2014). Sent home and put off-track: The antecedents, disproportionalities, and consequences of being suspended in the ninth grade. *Journal of Applied Research on Children: Informing Policy for Children at Risk*, *5*(2). Retrieved from `http://digitalcommons.library.tmc.edu/childrenatrisk/vol5/iss2/13`

Becker, G. S. (1993). *Human capital: A theoretical and empirical analysis with special reference to education* (3rd ed.). Chicago, IL: University of Chicago Press.

Bernburg, J. G., Krohn, M. D., & Rivera, C. J. (2006). Official labeling, criminal embeddness, and subsequent delinquency: A longitudinal test of labeling theory. *Journal of Research in Crime and Delinquency*, *43*(1), 67–88.

Bettinger, E. P., Terry Long, B., Oreopoulos, P., & Sanbonmatsu, L. (2012). The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment. *Quarterly Journal of Economics*, 1205–1242. doi: 10.1093/qje/qjs017.Advance

Blakemore, S.-J., & Choudhury, S. (2006). Development of the adolescent brain: Implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry*, *47*(3-4), 296–312. doi: 10.1111/j.1469-7610.2006.01611.x

Bol, L., & Berry, R. Q. (2005). Secondary mathematics teachers' perceptions of the achievement gap. *High School Journal*, *88*(4), 32–45. doi: 10.1353/hsj.2005.0007

Bourdieu, P. (1994). *Raisons pratiques: Sur la théorie de l'action*. Paris, France: Éditions du Seuil.

Bourdieu, P., & Passeron, J.-C. (1964). *Les héritiers: Les étudiants et la culture*. Paris, France: Les Éditions de Minuit.

Bowditch, C. (1993). Getting rid of troublemakers: High school disciplinary procedures and the production of dropouts. *Social Problems*, *40*(4), 493–509. doi: 10.3758/BF03334953

Bowers, A. J., Sprott, R., & Taff, S. A. (2012). Do we know who will drop out? A review of the predictors of dropping out of high school: Precision, sensitivity, and specificity. *High School Journal*, *96*(2), 77–100. Retrieved from http://www.jstor.org/stable/23351963

Bowman-Perrott, L., Benz, M. R., Hsu, H.-Y., Kwok, O.-M., Eisterhold, L. A., & Zhang, D. (2013). Patterns and predictors of disciplinary exclusion over time: An analysis of the SEELS national data set. *Journal of Emotional and Behavioral Disorders*, *21*(2), 83–96. doi: 10.1177/1063426611407501

Brand, J. E., & Xie, Y. (2010). Who benefits most from college? Evidence for negative selection in heterogeneous economic returns to higher education. *American Sociological Review*, *75*(2), 273–302. doi: 10.1177/0003122410363567

Brockenbrough, E. (2015). "The discipline stop": Black male teachers and the politics of urban school discipline. *Education and Urban Society*, *47*(5), 499–522. doi: 10.1177/0013124514530154

Bryc, K., Durand, E. Y., Macpherson, J. M., Reich, D., & Mountain, J. L. (2015). The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *American Journal of Human Genetics*, *96*(1), 37–53. doi: 10.1016/j.ajhg.2014.11.010

Buchmann, C., Condron, D. J., & Roscigno, V. J. (2010). Shadow education, American style: Test preparation, the SAT and college enrollment. *Social Forces*, *89*(2), 435–462. doi: 10.1353/sof.2010.0105

Bureau of Labor Statistics. (2016). *Earnings and unemployment rates by educational attainment, 2015.* Washington, D.C.. Retrieved from http://www.bls.gov/emp/ep_chart_001.htm

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, *69*(5), 1127–1160. doi: 10.1111/1468-0262.00237

Card, D., & Rothstein, J. (2007). Racial segregation and the black-white test score gap. *Journal of Public Economics*, *91*(11-12), 2158–2184. doi: 10.1016/j.jpubeco.2007.03.006

Caro, D. H., Cortina, K. S., & Eccles, J. S. (2015). Socioeconomic background, education, and labor force outcomes: Evidence from a regional US sample. *British Journal of Sociology of Education*, *36*(6), 934–957. doi: 10.1080/01425692.2013.868784

Carrell, S. E., & Sacerdote, B. (2013). *Why do college going interventions work?* Cambridge, MA. doi: 10.1017/CBO9781107415324.004

Carter, P. L. (2005). *Keepin' it real: School success beyond Black and White.* New York, NY: Oxford University Press.

Chein, J. M., Albert, D., O'Brien, L., Uckert, K., & Steinberg, L. (2011). Peers increase adolescent risk taking by enhancing activity in the brain's reward circuitry. *Developmental Science*, *14*(2), 1–10. doi: 10.1111/j.1467-7687.2010.01035.x

Children's Defense Fund. (1974). *Children out of school in America.* Cambridge, MA.

Children's Defense Fund. (1975). *School suspensions: Are they helping children?* Cambridge, MA. Retrieved from http://diglib.lib.utk.edu/cdf/data/0116_000050_000205/

0116_000050_000205.pdf

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. (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

Cobb-Clark, D. A., Kassenboehmer, S. C., Le, T., McVicar, D., & Zhang, R. (2015). Is there an educational penalty for being suspended from school? *Education Economics*, *23*(4), 376–395. doi: 10.1080/09645292.2014.980398

Columbia School District. (2012). *Regulations for student class work during suspension.* Columbia Station, OH. Retrieved from http://www.columbia.k12.oh.us/downloads/JGD-R6_20_12.pdf

Condron, D. J. (2009). Social class, school and non-school environments, and black/white inequalities in children's learning. *American Sociological Review*, *74*, 683–708.

Condron, D. J., & Roscigno, V. J. (2003). Disparities within: Unequal spending and achievement in an urban school district. *Sociology of Education*, *76*(1), 18–36. doi: 10.1126/science.135.3503.554

Czrnawski, H. C. (1954). *The construction of a student handbook for Ashland High School, Ashland, Massachusetts.* (Master's thesis, Boston University). Retrieved from https://open.bu.edu/bitstream/handle/2144/13021/Czrnawski_Helen_Catherine_1957_web.pdf?sequence=1

Dahl, R. E. (2004). Adolescent brain development: A period of vulnerabilities and opportunities. *Annals of the New York Academy of Sciences*, *1021*, 1–22. doi: 10.1196/annals.1308.011

Davis, P., Levine, R., & Part, S. (2015). *Civil rights suspended: An analysis of New York City charter school discipline policies.* New York, NY: Advocates for Children of New York. Retrieved from http://www.advocatesforchildren.org/sites/default/files/library/civil_rights_suspended.pdf?pt=1

DeSilver, D. (2014). *School days: How the U.S. compares with other countries.* Retrieved from http://www.pewresearch.org/fact-tank/2014/09/02/school-days-how-the-u-s-compares-with-other-countries/

Diamond, J. B., Randolph, A., & Spillane, J. P. (2004). Teachers' expectations and sense of responsibility for student learning: The importance of race, class, and organizational habitus. *Anthropology & Education Quarterly*, *35*(1), 75–98. doi: 10.1525/aeq.2004.35.1.75

Dominus, S. (2016). *An effective but exhausting alternative to high-school suspensions.* New York, NY. Retrieved from http://www.nytimes.com/2016/09/11/magazine/an-effective-ut-exhausting-alternative-to-high-school-suspensions.html?&moduleDetail=section-news-1&action=click&contentCollection=Magazine&region=Footer&module=MoreInSection&version=WhatsNext&contentID=WhatsNext&pgtype=article

Downey, D. B., von Hippel, P. T., & Broh, B. A. (2004). Are the schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review*, *69*, 613–635. doi: 10.1177/000312240406900501

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446. doi: 10.1037/0012-1649.43.6.1428

Duncan, G. J., & Magnuson, K. (2011). The nature and impact of early achievement skills, attention skills, and behavior problems. In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity? rising inequality, schools, and children's life chances* (pp. 47–69). New York, NY: Russell Sage Foundation.

Duncan, G. J., & Murnane, R. J. (2011). Introduction: The American dream, then and now. In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity? rising inequality, schools, and children's life chances* (pp. 3–23). New York, NY: Russell Sage Foundation.

Elfner Childs, K., Kincaid, D., Peshak George, H., & Gage, N. A. (2016). The relationship between school-wide implementation of positive behavior intervention and supports and student discipline outcomes. *Journal of Positive Behavior Interventions*, *18*(2), 89–99. doi: 10.1177/1098300715590398

Fabelo, T., Thompson, M. D., Plotkin, M., Carmichael, D., Marchbanks, III, M. P., & Booth, E. A. (2011). *Breaking schools' rules: A statewide study of how school discipline relates to students' success and juvenile justice involvement.* New York, NY: Council of State Governments Justice Center.

Feliciano, C., & Ashtiani, M. (2012). How low-income origins affect postsecondary entry and degree completion. *International Journal of Sociology of Education*, *1*(2), 123–156. doi: 10.4471/rise.2012.08

Figlio, D. N. (2006). Testing, crime, and punishment. *Journal of Public Economics*, *90*(4–5), 847–851. doi: 10.1016/j.jpubeco.2005.01.003

Freeman, J. B., Penner, A. M., Saperstein, A., Scheutz, M., & Ambady, N. (2011, jan). Looking the part: Social status cues shape race perception. *PLoS one*, *6*(9), e25107. doi: 10.1371/journal.pone.0025107

Fryer, R. G., Jr., & Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. *The Review of Economics and Statistics*, *86*(2), 447–464.

Gardiner, J. C., Luo, Z., & Roman, L. A. (2009). Fixed effects, random effects and GEE: What are the differences? *Statistics in Medicine*, *28*, 221–239. doi: 10.1002/sim.3478

Gardner, M., & Steinberg, L. (2005). Peer influence on risk taking, risk preference, and risky decision making in adolescence and adulthood: An experimental study. *Developmental Psychology*, *41*(4), 625–635. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/16060809` doi: 10.1037/0012-1649.41.4.625

GED Testing Service. (2014). *2013 Annual statistical report on the GED test.* Retrieved from `http://www.gedtestingservice.com/uploads/files/5b49fc887db0c075da20a68b17d313cd.pdf`

Glassman, L. A. (1943). Is dull normal intelligence a contraindication for psychotherapy? *Smith College Studies in Social Work*, *13*(3), 275–298. doi: 10.1080/00377314309512800

Goldin, C., Katz, L. F., & Kuziemko, I. (2006). The homecoming of American college women: The reversal of the college gender gap. *Journal of Economic Perspectives*, *20*(4), 133–156. doi: 10.1257/jep.20.4.133

Gordon, L. (2007). Does a pot bust trump a 4.0 GPA? *Los Angeles Times*. Retrieved from `http://www.latimes.com/local/la-me-admit5dec05-story.html`

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

455–475. doi: 10.1016/j.jsp.2007.09.001

Grodsky, E., & Jones, M. T. (2007). Real and imagined barriers to college entry: Perceptions of cost. *Social Science Research*, *36*(2), 745–766. doi: 10.1016/j.ssresearch.2006.05.001

Grodsky, E., & Riegle-Crumb, C. (2010). Those who choose and those who don't: Social background and college orientation. *The ANNALS of the American Academy of Political and Social Science*, *627*(1), 14–35. doi: 10.1177/0002716209348732

Haney López, I. (2006). *White by law: The legal construction of race* (10th anniversary ed.). New York, NY: New York University Press.

Harris, A. L. (2006). I (don't) hate school: Revisiting oppositional culture theory of blacks' resistance to schooling. *Social Forces*, *85*(2), 797–834. doi: 10.1353/sof.2007.0006

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.

Hayes, D. (2011). Predicting parental home and school involvement in high school African American adolescents. *High School Journal*, *94*(4), 154–166. doi: 10.1353/hsj.2011.0010

Heckman, J. J., Humphries, J. E., & Mader, N. S. (2010). *The GED.* doi: 10.1016/B978-0-444-53429-3.00009-0

Heckman, J. J., Ichimura, H., & Todd, P. (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies*, 261–294. doi: 10.1111/1467-937X.00044

Heckman, J. J., & LaFontaine, P. A. (2010). The American high school graduation rate: Trends and levels. *Review of Economics and Statistics*, *92*(2), 244–262. doi: 10.1162/rest.2010.12366

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, *29*(1), 60–87. doi: 10.3102/0162373707299706

Hill, L. D. (2008). School strategies and the "college-linking" process: Reconsidering the effects of high schools on college enrollment. *Sociology of Education*, *81*(1), 53–76. doi: 10.1177/003804070808100103

Hirschfield, P. J. (2008). Preparing for prison?: The criminalization of school discipline in the USA. *Theoretical Criminology*, *12*(1), 79–101. doi: 10.1177/1362480607085795

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*(396), 945–960. doi: 10.2307/2289064

Hong, G., & Raudenbush, S. W. (2003). Causal inference for multi-level observational data with application to kindergarten retention study. In *2003 joint statistical meetings proceedings* (pp. 1849–1856). Alexandria, VA: American Statistical Association.

Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A. W., & Esperanza, J. (2009). A randomized, wait-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior Interventions*, *11*(3), 133–144. doi: 10.1177/1098300709332067

Horner, S. B., Fireman, G. D., & Wang, E. W. (2010). The relation of student behavior, peer status, race, and gender to decisions about school discipline using CHAID decision trees and regression modeling. *Journal of School Psychology*, *48*(2), 135–161. doi: 10.1016/j.jsp.2009.12.001

Hossler, D., & Gallagher, K. (1987). Studying college choice: A three-phase model and the implications for policy makers. *College and University*, *2*, 207–221.

Hout, M. (2012). Social and economic returns to college education in the United States. *Annual Review of Sociology*, *38*(1), 379–400. doi: 10.1146/annurev.soc.012809.102503

Iacus, S. M., King, G., & Porro, G. (2009). CEM: Coarsened exact matching software. *Journal of Statistical Software*, *30*(9), 1–27.

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, *20*(1), 1–24. doi: 10.1093/pan/mpr013

Jackson, C. K. (2009). Student demographics, teacher sorting, and teacher quality: Evidence from the end of school desegregation. *Journal of Labor Economics*, *27*(2), 213–256. doi: 10.1086/599334

Jenkins, D., & Fink, J. (2016). *Tracking transfer: New measures of institutional and state effectiveness in helping community college students attain bachelor's degrees.* New York, NY: Community College Research Center, Teachers College, Columbia University. Retrieved from `http://ccrc.tc.columbia.edu/media/k2/attachments/tracking-transfer-institutional-state-effectiveness.pdf`

Jimerson, S. R., Anderson, G. E., & Whipple, A. D. (2002). Winning the battle and losing the war: Examining the relation between grade retention and dropping out of high school. *Psychology in the Schools*, *39*(4), 441–457. doi: 10.1002/pits.10046

Kafka, J. (2011). *The history of "zero tolerance" in American public schooling.* New York, NY: Palgrave Macmillan.

Kann, L., Kinchen, S., Shanklin, S. L., Flint, K. H., Hawkins, J., Harris, W. A., ... Zaza, S. (2014). *Youth risk behavior surveillance - United States, 2013.* Atlanta, GA: Centers for Disease Control and Prevention, U.S. Department of Health and Human Services. Retrieved from `http://www.cdc.gov/mmwr/pdf/ss/ss6304.pdf`

Keele, L. (2010). *An overview of rbounds: An R package for Rosenbaum bounds sensitivity analysis with matched data.* Retrieved from `http://www.personal.psu.edu/ljk20/rboundsvignette.pdf`

Keele, L. (2015). The statistics of causal inference: A view from political methodology. *Political Analysis*, *23*(3), 313–335. doi: 10.1093/pan/mpv007

Kena, G., Musu-Gillette, L., Robinson, J., Wang, X., Rathbun, A., Zhang, J., ... Dunlop Velez, E. (2015). *The condition of education 2015 (NCES 2015-144).* Washington, D.C.: U.S. Department of Education, National Center for Education Statistics. Retrieved from `http://nces.ed.gov/pubs2015/2015144.pdf`

Kim, C. Y., Losen, D. J., & Hewitt, D. T. (2010). *The school-to-prison pipeline: Structuring legal reform.* New York, NY: New York University Press.

King, G., & Nielsen, R. (n.d.). *Why propensity scores should not be used for matching.* Retrieved from `https://gking.harvard.edu/publications/why-propensity-scores-should-not-be-used-formatching`

Kinsler, J. (2011). Understanding the black-white school discipline gap. *Economics of Education Review*, *30*(6), 1370–1383. doi: 10.1016/j.econedurev.2011.07.004

Kinsler, J. (2013). School discipline: A source or salve for the racial achievement gap? *International Economic Review*, *54*(1), 355–383. doi: 10.1111/j.1468-2354.2012.00736.x

Kunesh, C. E., & Noltemeyer, A. (2015). Understanding disciplinary disproportionality: Stereotypes shape pre-service teachers' beliefs about Black boys' behavior. *Urban Education*, *Advance online publication*. doi: 10.1177/0042085915623337

Kupchik, A. (2010). *Homeroom security: School discipline in an age of fear*. New York, NY: New York University Press.

Lareau, A. (2011). *Unequal childhoods: Class, race, and family life* (2nd ed.). Berkeley, CA: University of California Press.

Lareau, A., & Horvat, E. M. (1999). Moments of social inclusion and exclusion: Race, class, and cultural capital in family-school relationships. *Sociology of Education*, *72*(1), 37–53.

Learned, J. E. (2016). "The behavior kids": Examining the conflation of youth reading difficulty and behavior problem positioning among school institutional contexts. *American Educational Research Journal*, *53*(5), 1271–1309. doi: 10.3102/0002831216667545

Lee, J., & Bean, F. D. (2004). America's changing color lines: Immigration, race/ethnicity, and multiracial identification. *Annual Review of Sociology*, *30*(1), 221–242. doi: 10.1146/annurev.soc.30.012703.110519

Lee, T., Cornell, D., Gregory, A., & Fan, X. (2011). High suspension schools and dropout rates for Black and White students. *Education and Treatment of Children*, *34*(2), 167–192. doi: 10.1353/etc.2011.0014

Levin, H. M. (2009). The economic payoff to investing in educational justice. *Educational Researcher*, *38*(1), 5–20. doi: 10.3102/0013189X08331192

Lewis, A. E., & Diamond, J. B. (2015). *Despite the best intentions: How racial inequality thrives in good schools*. New York, NY: Oxford University Press.

Li, C. (2016). *The effectiveness and consequences of exclusionary school discipline*. Retrieved from `http://chunxiao.web.unc.edu/files/2016/11/Li2016.pdf`

Losen, D., Hodson, C., Keith, II, M. A., Morrison, K., & Belway, S. (2015). *Are we closing the school discipline gap?* Los Angeles, CA: The Center for Civil Rights Remedies at The Civil Rights Project, UCLA. Retrieved from `https://www.civilrightsproject.ucla.edu/resources/projects/center-for-civil-rights-remedies/school-to-prison-folder/federal-reports/are-we-closing-the-school-discipline-gap/losen-are-we-closing-discipline-gap-2015-summary.pdf`

Losen, D. J. (2011). *Discipline policies, successful schools, and racial justice*. Boulder, CO: National Education Policy Center.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *American Economic Review*, *80*(2), 319–323. Retrieved from `http://www.jstor.org/stable/pdf/2006592.pdf`

Manski, C. F. (1997). Monotone treatment response. *Econometrica*, *65*(6), 1311–1334. Retrieved from `http://www.jstor.org/stable/2171738`

Manski, C. F., & Pepper, J. V. (2000). Monotone instrumental variables: With an application to the returns to schooling. *Econometrica*, *68*(4), 997–1010. Retrieved from `http://onlinelibrary.wiley.com/doi/10.1111/1468-0262.t01-1-00144a/epdf`

McCarthy, J. D., & Hoge, D. R. (1987). The social construction of school punishment: Racial disadvantage out of universalistic process. *Social Forces*, *65*(4), 1101–1120.

Mickelson, R. A. (2015). The cumulative disadvantages of first- and second-generation segregation for middle school achievement. *American Educational Research Journal*, *52*(4), 657–692. doi: 10.3102/0002831215587933

Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference: Methods and principles for social research* (2nd ed.). New York, NY: Cambridge University Press.

Morris, E. W., & Perry, B. L. (2016). The punishment gap: School suspension and racial disparities in achievement. *Social Problems*, *63*(1), 68–86. doi: 10.1093/socpro/spv026

Newman, L., Wagner, M., Cameto, R., Knokey, A.-M., & Shaver, D. (2010). *Comparisons across time of the outcomes of youth with disabilities up to 4 years after high school: A report of the findings from the National Longitudinal Transition Study (NLTS) and the National Longitudinal Transition Study-2 (NLTS2).* Washington, D.C.: Institute of Education Sciences. Retrieved from `http://files.eric.ed.gov/fulltext/ED512149.pdf`

Nicholson-Crotty, S., Birchmeier, Z., & Valentine, D. (2009). Exploring the impact of school discipline on racial disproportion in the juvenile justice system. *Social Science Quarterly*, *90*(4), 1003–1018. doi: 10.1111/j.1540-6237.2009.00674.x

Noltemeyer, A. L., Ward, R. M., & Mcloughlin, C. (2015). Relationship between school suspension and student outcomes: A meta-analysis. *School Psychology Review*, *44*(2), 224–240. doi: 10.17105/spr-14-0008.1

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., ... Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*(1), 36–88. doi: 10.1080/10463280701489053

Ogbu, J. U. (1987). Variability in minority school performance: A problem in search of an explanation. *Anthropology & Education Quarterly*, *18*, 312–334. Retrieved from `http://www.jstor.org/stable/3216660`

Okonofua, J. A., & Eberhardt, J. L. (2015). Two strikes: Race and the disciplining of young students. *Psychological Science*, *26*(5), 617–624. doi: 10.1177/0956797615570365

Omi, M., & Winant, H. (2014). *Racial formation in the United States* (3rd ed.). New York, NY: Routledge.

Orfield, G., Ee, J., Frankenberg, E., & Siegel-Hawley, G. (2016). *Brown At 62: School segregation by race, poverty and state.* Los Angeles, CA: Civil Rights Project, UCLA. Retrieved from `https://www.civilrightsproject.ucla.edu/research/k-12-education/integration-and-diversity/brown-at-62-school-segregation-by-race-poverty-and-state/Brown-at-62-final-corrected-2.pdf`

Pager, D. (2003). The mark of a criminal record. *American Journal of Sociology*, *108*(5), 937–975. Retrieved from `http://www.jstor.org/stable/10.1086/374403`

Pager, D., & Shepherd, H. (2008). The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. *Annual Review of Sociology*, *34*, 181–209. doi: 10.1146/annurev.soc.33.040406.131740

Pager, D., Western, B., & Bonikowski, B. (2009). Discrimination in a low-wage labor market: A field experiment. *American Sociological Review*, *74*, 777–799. doi: 10.1177/000312240907400505

Pallais, A. (2015). Small differences that matter: Mistakes in applying to college. *Journal of Labor Economics*, *33*(2), 493–520. doi: 10.1086/678520

Perna, L. W. (2006). Studying college access and choice: A proposed conceptual model. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. XXI, pp. 99–157). Springer. Retrieved from `http://link.springer.com/chapter/10.1007/`

1-4020-4512-3_3#page-1

Perna, L. W., & Titus, M. A. (2005). The relationship between parental involvement as social capital and college enrollment: An examination of racial/ethnic group differences. *Journal of Higher Education*, *76*(5), 485–518. doi: 10.1353/jhe.2005.0036

Perry, B. L., & Morris, E. W. (2014). Suspending progress: Collateral consequences of exclusionary punishment in public schools. *American Sociological Review*, *79*(6), 1–21. doi: 10.1177/0003122414556308

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata* (3rd ed.). College Station, Texas: Stata Press.

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. Retrieved from www.jstor.org/stable/42900535

Rastogi, S., Johnson, T. D., Hoeffel, E. M., & Drewery, M. P. (2011). *The Black population: 2010*. Washington, D.C.: United States Census Bureau. Retrieved from https://www.census.gov/prod/cen2010/briefs/c2010br-06.pdf

Reardon, S. F., & Owens, A. (2014). 60 years after Brown: Trends and consequences of school segregation. *Annual Review of Sociology*, *40*(1), 199–218. doi: 10.1146/annurev-soc-071913-043152

Rickles, J. H., & Seltzer, M. (2014). A two-stage propensity score matching strategy for treatment effect estimation in a multisite observational study. *Journal of Educational and Behavioral Statistics*, *39*(6), 612–636. doi: 10.3102/1076998614559748

Robinson, K., & Harris, A. L. (2014). *The broken compass: Parental involvement with children's education*. Cambridge, MA: Harvard University Press.

Roderick, M., Coca, V., & Nagaoka, J. (2011). Potholes on the road to college: High school effects in shaping urban students' participation in college application, four-year college enrollment, and college match. *Sociology of Education*, *84*(3), 178–211. doi: 10.1177/0038040711411280

Rosenbaum, P. R. (2002). *Observational studies* (2nd ed.). New York, NY: Springer.

Rosenbaum, P. R. (2010). *Design of observational studies*. New York, NY: Springer. doi: 10.1007/978-0-387-98135-2

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701. doi: 10.1037/h0037350

Rubin, D. B. (2005). Causal inference using potential outcomes. *Journal of the American Statistical Association*, *100*(469), 322–331. doi: 10.1198/016214504000001880

Rumberger, R. W., & Losen, D. J. (2016). *The high cost of harsh discipline and its disparate impact*. Los Angeles, CA: The Center for Civil Rights Remedies at The Civil Rights Project. Retrieved from https://www.civilrightsproject.ucla.edu/resources/projects/center-for-civil-rights-remedies/school-to-prison-folder/federal-reports/the-high-cost-of-harsh-discipline-and-its-disparate-impact/UCLA_HighCost_6-2_948.pdf

Ryan, C. L., & Bauman, K. (2016). *Educational attainment in the United States*. Washington, D.C.: U.S. Census Bureau. Retrieved from http://www.census.gov/prod/2012pubs/p20-566.pdf

Sanders, S. G., & Mendez, R. (1981). An examination of in-school suspension: Panacea or Pandora's box? *NASSP Bulletin*, *65*, 65–69. Retrieved from http://journals.sagepub.com/doi/abs/10.1177/019263658106544114?journalCode=buld

Saperstein, A., & Penner, A. M. (2012). Racial fluidity and inequality in the United States. *American Journal of Sociology*, *118*(3), 676–727. doi: 10.1163/_afco_asc_2291

Schneider, B., & Saw, G. (2016). Racial and ethnic gaps in postsecondary aspirations and enrollment. *Russell Sage Foundation Journal of the Social Sciences*, *2*(5), 58–82. Retrieved from https://muse.jhu.edu/article/633737/pdf

Seeley, L. (1901). *The foundations of education*. New York, NY: Hinds & Noble. Retrieved from https://catalog.hathitrust.org/Record/006510576

Sekhon, J. S. (2011). Multivariate and propensity score matching software with automated balance optimization: The Matching package for R. *Journal of Statistical Software*, *42*(7), 1–52. Retrieved from http://www.jstatsoft.org/v42/i07/

Skiba, R. J., Chung, C.-G., 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. doi: 10.3102/0002831214541670

Skiba, R. J., Horner, R. H., Chung, C.-G., Rausch, M. K., May, S. L., & Tobin, T. (2011). Race is not neutral: A national investigation of African American and Latino disproportionality in school discipline. *School Psychology Review*, *40*(1), 85–107. Retrieved from http://www.indiana.edu/~equity/docs/Skiba%20et%20al%20Race%20is%20Not%20Neutral%202011.pdf

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. *The Urban Review*, *34*(4), 317–342. doi: 10.1023/A:1021320817372

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Thousand Oaks, CA: Sage.

Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. *Developmental Review*, *28*(1), 78–106. doi: 10.1016/j.dr.2007.08.002

Steiner, P. M., Kim, J.-S., & Thoemmes, F. (2012). Matching strategies for observational multilevel data. In *Joint statistical meetings proceedings* (pp. 5020–5032). San Diego, CA. Retrieved from http://itp.wceruw.org/documents/steiner_kim_thoemmes_JSM13.pdf

Stewart, E. A. (2003). School social bonds, school climate, and school misbehavior: A multilevel analysis. *Justice Quarterly*, *20*(3), 575–604. doi: 10.1080/07418820300095621

Strulik, H., Prettner, K., & Prskawetz, A. (2013). The past and future of knowledge-based growth. *Journal of Economic Growth*, *18*(4), 411–437. doi: 10.1007/s10887-013-9098-9

Sullivan, A. L., Klingbeil, D. A., & Van Norman, E. R. (2013). Beyond behavior: Multilevel analysis of the influence of sociodemographics and school charateristics on students' risk of suspension. *School Psychology Review*, *42*(1), 99–114. Retrieved from https://search.proquest.com/openview/fb33a4f43b32623eba2fb1675922b7d4/1?pq-origsite=gscholar&cbl=48217

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

Educational Psychology, *99*(2), 253–273. doi: 10.1037/0022-0663.99.2.253

Theriot, M. T., Craun, S. W., & Dupper, D. R. (2010). Multilevel evaluation of factors predicting school exclusion among middle and high school students. *Children and Youth Services Review*, *32*(1), 13–19. doi: 10.1016/j.childyouth.2009.06.009

Tobin, T., Sugai, G., & Colvin, G. (1996). Patterns in middle school discipline records. *Journal of Emotional and Behavioral Disorders*, *4*(2), 82–94. doi: 10.1177/106342669600400203

U.S. Department of Education. (n.d.). *CRDC 2011-12 Definitions.* `http://ocrdata.ed.gov/DataDefinitions`. (Accessed: 2016-05-09)

U.S. Department of Education Office for Civil Rights [OCR]. (2014). *Civil rights data collection: Data snapshot: School discipline (Issue Brief No. 1).* Retrieved from `http://www2.ed.gov/about/offices/list/ocr/docs/crdc-discipline-snapshot.pdf`

van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. W. (2010). The implicit prejudiced attitudes of teachers: Relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal*, *47*(2), 497–527. doi: 10.3102/0002831209353594

Vanneman, A., Hamilton, L., Anderson, J. B., & Rahman, T. (2009). *Achievement gaps: How black and white students in public schools perform in mathematics and reading on the National Assessment of Educational Progress (NCES 2009-455).* Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Vavrus, F., & Cole, K. (2002). "I didn't do nothin'": The discursive construction of school suspension. *Urban Review*, *34*(2), 87–111. doi: 10.1023/A:1015375215801

Wallace, J. M., Jr., Goodkind, S., Wallace, C. M., & Bachman, J. G. (2008). Racial, ethnic, and gender differences in school discipline among U.S. high school students: 1991-2005. *The Negro Educational Review*, *59*(1-2), 47–62. Retrieved from `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2678799/`

Weinstein, R. S. (2002). *Reaching higher: The power of expectations in schooling.* Cambridge, MA: Harvard University Press.

Weissman, M., & NaPier, E. (2015). *Education suspended: The use of high school disciplinary records in college admissions.* Syracuse, NY: Center for Community Alternatives. Retrieved from `http://www.communityalternatives.org/pdf/publications/EducationSuspended.pdf`

Zhang, A., Musu-Gillette, L., & Oudekerk, B. A. (2016). *Indicators of school crime and safety: 2015.* Washington, D.C.: U.S. Department of Education. Retrieved from `http://eric.ed.gov/ERICWebPortal/recordDetail?accno=EJ646488` doi: 10.1037/e541412012-001

Table 6.1: Descriptive Statistics for the Full Sample

Variable	Proportion (%)	SE		
Suspension	13.51	0.33		
American Indian/Alaska Native	0.98	0.09		
Asian/Pacific Islander	11.11	0.30		
Black	14.12	0.34		
Hispanic	15.23	0.35		
More than one race	4.87	0.21		
White	53.68	0.48		
Male	48.37	0.48		
Participated in sports	51.15	0.49		
History of retention	12.59	0.36		
IEP status: Yes	6.69	0.24		
IEP status: No	45.80	0.48		
IEP status: Missing	47.50	0.48		
Single-parent household	23.89	0.41		
Parent college graduate	35.82	0.46		
Northeast	16.70	0.36		
Midwest	24.27	0.41		
South	37.73	0.47		
West	21.30	0.39		
Urban	26.72	0.43		
Suburban	50.78	0.48		
Rural	22.49	0.40		
	Mean	SD	Min	Max
Math and reading achievement	-0.10	9.92	-28.5	29.94
Household socioeconomic status	-0.07	0.71	-2.11	1.98
School-level achievement	-0.16	5.12	-15.23	17.35
School-level socioeconomic status	-0.06	0.36	-1.07	1.19

$(N = 10,799)$

Table 6.2: Descriptive Statistics for the Sample Not Missing Covariate Data

Variable	Proportion (%)	SE		
Suspension	12.28	0.36		
American Indian/Alaska Native	0.89	0.10		
Asian/Pacific Islander	9.56	0.32		
Black	12.65	0.36		
Hispanic	15.26	0.39		
More than one race	4.60	0.23		
White	57.04	0.54		
Male	47.84	0.54		
Participated in sports	52.77	0.54		
History of retention	12.32	0.36		
IEP status: Yes	6.02	0.26		
IEP status: No	47.46	0.54		
IEP status: Missing	46.52	0.54		
Single-parent household	23.07	0.46		
Parent college graduate	37.15	0.52		
Northeast	16.78	0.41		
Midwest	24.21	0.46		
South	37.88	0.53		
West	21.14	0.44		
Urban	25.04	0.47		
Suburban	51.32	0.54		
Rural	23.64	0.46		
	Mean	SD	Min	Max
Math and reading achievement	0.74	9.84	-28.5	29.94
Household socioeconomic status	-0.02	0.72	-1.97	1.98
School-level achievement	0.18	4.76	-15.23	17.35
School-level socioeconomic status	-0.05	0.36	-1.07	1.19

$(N = 8,506)$

Table 6.3: No-Assumptions Bounds

	$E[Y(1) \mid .]$	$E[Y(0) \mid .]$
Naive estimator suggests $E[\delta] =$		
Treatment group	$E[Y(1) \mid D = 1] = 0.20$	$E[Y(0) \mid D = 1] =?$
Control group	$E[Y(1) \mid D = 0] =?$	$E[Y(0) \mid D = 0] = 0.47$
Largest possible $E[\delta] = 0.48$		
Treatment group	$E[Y(1) \mid D = 1] = 0.20$	$E[Y(0) \mid D = 1] = 0$
Control group	$E[Y(1) \mid D = 0] = 1$	$E[Y(0) \mid D = 0] = 0.47$
Smallest possible $E[\delta] = -0.52$		
Treatment group	$E[Y(1) \mid D = 1] = 0.20$	$E[Y(0) \mid D = 1] = 1$
Control group	$E[Y(1) \mid D = 0] = 0$	$E[Y(0) \mid D = 0] = 0.47$

Table 6.4: Bounds Under Monotonic Treatment Selection and Response

	$E[Y(1) \mid .]$	$E[Y(0) \mid .]$
Largest possible $E[\delta] = 0$		
Treatment group	$E[Y(1) \mid D = 1] = 0.20$	$E[Y(0) \mid D = 1] = 0.20$
Control group	$E[Y(1) \mid D = 0] = 0.47$	$E[Y(0) \mid D = 0] = 0.47$
Smallest possible $E[\delta] = -0.27$		
Treatment group	$E[Y(1) \mid D = 1] = 0.20$	$E[Y(0) \mid D = 1] = 0.47$
Control group	$E[Y(1) \mid D = 0] = 0.20$	$E[Y(0) \mid D = 0] = 0.47$

Table 6.5: Estimated Odds Ratios and 95% Confidence Intervals for Suspension

Fixed part	Odds Ratio	95% Confidence Interval
American Indian	1.044	[0.546, 1.997]
Asian/Pacific Islander	0.496*	[0.346, 0.713]
Black	1.586*	[1.282, 1.962]
Hispanic	0.972	[0.744, 1.270]
More than one race	1.410*	[1.023, 1.944]
Male	1.609*	[1.395, 1.855]
Participated in sports	0.742*	[0.646, 0.852]
Math and reading achievement	0.950*	[0.941, 0.960]
History of retention	1.467*	[1.236, 1.743]
IEP status: No	0.818	[0.618, 1.082]
IEP status: Missing	0.786	[0.599, 1.031]
Single-parent household	1.363*	[1.161, 1.600]
Parent college graduate	0.830	[0.668, 1.030]
Household SES	0.912	[0.779, 1.067]
School mean achievement	1.003	[0.976, 1.030]
School mean SES	0.859	[0.595, 1.239]
Northeast	1.149	[0.889, 1.484]
Midwest	1.130	[0.878, 1.455]
South	1.366*	[1.075, 1.736]
Urban	1.068	[0.874, 1.304]
Rural	0.970	[0.786, 1.198]
Constant	0.075	
Random part		
Random-intercept variance	0.164	
Intraclass correlation	0.047	
N observations	8,506	
N schools	578	

Note. A * indicates significant at $p < 0.05$ (excluding the intercept). White is the reference category for race/ethnicity. Having an IEP is the reference category for IEP status. West is the reference category for region. Suburban is the reference category for school locale. School size covariates were included in the model but are omitted from the table.

Table 6.6: Estimated Odds Ratios and 95% Confidence Intervals for Matriculation at a Four-Year College

Fixed part	Model 1	Model 2	Fixed Effects
Suspension	0.457* [0.361, 0.578]	0.454* [0.359, 0.576]	0.448* [0.361, 0.555]
American Indian	1.731 [0.952, 3.147]	1.857* [1.043, 3.306]	2.380* [1.135, 4.989]
Asian/Pacific Islander	2.278* [1.806, 2.873]	2.412* [1.896, 3.068]	2.317* [1.773, 3.027]
Black	2.503* [2.029, 3.087]	2.429* [1.961, 3.008]	2.137* [1.683, 2.713]
Hispanic	1.066 [0.858, 1.324]	1.170 [0.936, 1.463]	1.024 [0.805, 1.303]
More than one race	1.052 [0.807, 1.373]	1.106 [0.848, 1.443]	1.047 [0.778, 1.410]
Male	0.640* [0.571, 0.717]	0.651* [0.581, 0.730]	0.644* [0.569, 0.728]
Participated in sport(s)	2.136* [1.876, 2.431]	2.143* [1.883, 2.439]	2.172 [1.913, 2.465]
Math and reading achievement	1.121* [1.110, 1.131]	1.118* [1.107, 1.129]	1.116* [1.106, 1.126]
History of retention	0.560* [0.448, 0.701]	0.549* [0.439, 0.687]	0.542* [0.432, 0.681]
IEP status: No	1.629* [1.079, 2.460]	1.688* [1.120, 2.544]	1.878* [1.182, 2.983]
IEP status: Missing	1.725* [1.146, 2.598]	1.820* [1.207, 2.746]	1.664* [1.044, 2.650]
Single-parent household	0.955 [0.818, 1.113]	0.933 [0.800, 1.088]	0.949 [0.816, 1.102]
Parent college graduate	1.545* [1.314, 1.816]	1.545* [1.315, 1.817]	1.480* [1.247, 1.757]
Household SES	1.636* [1.438, 1.861]	1.540* [1.351, 1.756]	1.565* [1.364, 1.795]
School mean achievement		1.009 [0.983, 1.035]	-
School mean SES		1.594* [1.137, 2.236]	-
Northeast		1.726* [1.349, 2.208]	-
Midwest		1.637* [1.316, 2.036]	-
South		1.580* [1.277, 1.956]	-
Urban		1.521* [1.251, 1.850]	-
Rural		0.999 [0.816, 1.222]	-
Constant	0.383	0.163	-
Random part			
Random-intercept variance	0.378	0.296	-
Intraclass correlation	0.103	0.083	-
Log-likelihood	-3,793.74	-3,764.48	-2,629.09
N observations	7,626	7,626	7,387
N schools	577	577	548

Note. A * indicates significant at $p < 0.05$ (excluding the intercept). Twenty-nine schools were excluded from the fixed effects analysis because all students either attended or did not attend college.

Table 6.7: Balance Before and After Propensity Score Matching Across Schools (Matched Variables)

Variable	Mean Treatment	Mean Control		t-test p-values		KS-test p-values	
		Before	After	Before	After	Before	After
American Indian	0.01	0.01	0.01	0.47	0.78		
Asian	0.04	0.10	0.03	0.00	0.23		
Black	0.25	0.11	0.27	0.00	0.18		
Latino	0.19	0.14	0.18	0.00	0.58		
Multiracial	0.06	0.04	0.06	0.09	0.87		
Male	0.55	0.45	0.59	0.00	0.05		
Sports	0.44	0.54	0.42	0.00	0.28		
Math/Reading	-4.63	1.92	-5.55	0.00	0.00	0.00	0.00
Retain	0.24	0.10	0.26	0.00	0.26		
SES	-0.26	0.03	-0.31	0.00	0.08	0.00	0.00
Single Parent	0.34	0.21	0.35	0.00	0.52		
IEP: Yes	0.12	0.05	0.13	0.00	0.47		
IEP: Missing	0.44	0.47	0.45	0.20	0.74		
Parent College	0.26	0.39	0.23	0.00	0.07		
Urban	0.31	0.24	0.30	0.00	0.86		
Rural	0.22	0.24	0.23	0.36	0.82		
Northeast	0.14	0.17	0.15	0.02	0.79		
Midwest	0.21	0.25	0.19	0.01	0.39		
South	0.47	0.37	0.50	0.00	0.24		
School Math/Reading	-1.43	0.51	-1.62	0.00	0.35	0.00	0.01
School SES	-0.14	-0.03	-0.16	0.00	0.30	0.00	0.00

Table 6.8: Balance Before and After Propensity Score Matching Across Schools (Other Variables)

Variable	Mean Treatment	Mean Control		t-test p-values	
		Before	After	Before	After
Native English Speaker	0.85	0.83	0.83	0.06	0.27
First-Generation Immigrant	0.08	0.11	0.10	0.00	0.14
Second-Generation Immigrant	0.10	0.13	0.10	0.01	0.74
Wealth	0.19	0.26	0.18	0.00	0.49
AP Class	0.12	0.21	0.13	0.00	0.39
IB Program	0.02	0.02	0.01	0.77	0.50
Risk Factors	1.50	0.94	1.47	0.00	0.60
Parent Native English Speaker	0.85	0.81	0.84	0.00	0.33
School College Matriculation Rate	0.38	0.45	0.39	0.00	0.88
School Suspension Rate	0.21	0.12	0.16	0.00	0.00

Table 6.9: Balance Before and After Propensity Score Matching Within Schools (Matched Variables)

Variable	Mean Treatment	Mean Control		t-test p-values		KS-test p-values	
		Before	After	Before	After	Before	After
American Indian	0.01	0.01	0.01	0.47	0.44		
Asian	0.04	0.10	0.04	0.00	0.90		
Latino	0.19	0.14	0.20	0.00	0.82		
Multiracial	0.06	0.04	0.06	0.09	1.00		
Male	0.55	0.45	0.52	0.00	0.25		
Sports	0.44	0.54	0.48	0.00	0.11		
Math/Reading	-4.63	1.92	-3.97	0.00	0.01	0.00	0.12
Retain	0.24	0.10	0.16	0.00	0.00		
SES	-0.26	0.03	-0.27	0.00	0.64	0.00	0.47
Single Parent	0.34	0.21	0.30	0.00	0.10		
IEP: Yes	0.12	0.05	0.09	0.00	0.02		
IEP: Missing	0.44	0.47	0.46	0.20	0.18		
Parent College	0.26	0.39	0.22	0.00	0.02		
Urban	0.31	0.24	0.31	0.00	1.00		
Rural	0.22	0.24	0.22	0.36	1.00		
Northeast	0.14	0.17	0.14	0.02	1.00		
Midwest	0.21	0.25	0.21	0.01	1.00		
South	0.47	0.37	0.47	0.00	1.00		
School Math/Reading	-1.43	0.51	-1.43	0.00	1.00	0.00	1.00
School SES	-0.14	-0.03	-0.14	0.00	1.00	0.00	1.00

Table 6.10: Balance Before and After Propensity Score Matching Within Schools (Other Variables)

Variable	Mean Treatment	Mean Control		t-test p-values	
		Before	After	Before	After
Native English Speaker	0.85	0.83	0.78	0.06	0.00
First-Generation Immigrant	0.08	0.11	0.10	0.00	0.05
Second-Generation Immigrant	0.10	0.13	0.13	0.01	0.02
Wealth	0.19	0.26	0.21	0.00	0.24
AP Class	0.12	0.21	0.14	0.00	0.16
IB Program	0.02	0.02	0.02	0.77	0.47
Risk Factors	1.50	0.94	1.28	0.00	0.00
Parent Native English Speaker	0.85	0.81	0.81	0.00	0.01
School College Matriculation Rate	0.38	0.45	0.38	0.00	1.00
School Suspension Rate	0.21	0.12	0.21	0.00	1.00

Table 6.11: Balance Before and After Genetic Matching Across Schools (Matched Variables)

Variable	Mean Treatment	Mean Control		t-test p-values		KS-test p-values	
		Before	After	Before	After	Before	After
American Indian	0.01	0.01	0.02	0.47	0.39		
Asian	0.04	0.10	0.05	0.00	0.56		
Black	0.25	0.11	0.25	0.00	1.00		
Latino	0.19	0.14	0.16	0.00	0.02		
Multiracial	0.06	0.04	0.04	0.09	0.14		
Male	0.55	0.45	0.53	0.00	0.26		
Sports	0.44	0.54	0.44	0.00	0.81		
Math/Reading	-4.63	1.92	-4.41	0.00	0.23	0.00	0.30
Retain	0.24	0.10	0.24	0.00	1.00		
SES	-0.26	0.03	-0.26	0.00	0.96	0.00	0.92
Single Parent	0.34	0.21	0.34	0.00	0.32		
IEP: Yes	0.44	0.48	0.44	0.01	0.32		
IEP: Missing	0.44	0.47	0.45	0.20	0.32		
Parent College	0.26	0.39	0.26	0.00	0.32		
Urban	0.31	0.24	0.31	0.00	1.00		
Rural	0.22	0.24	0.22	0.36	0.41		
Northeast	0.14	0.17	0.13	0.02	0.31		
Midwest	0.21	0.25	0.21	0.01	0.32		
South	0.47	0.37	0.48	0.00	0.25		
School Math/Reading	-1.43	0.51	-1.36	0.00	0.24	0.00	0.75
School SES	-0.14	-0.03	-0.14	0.00	0.96	0.00	0.99

Table 6.12: Balance Before and After Genetic Matching Across Schools (Other Variables)

Variable	Mean Treatment	Mean Control		t-test p-values	
		Before	After	Before	After
Native English Speaker	0.85	0.83	0.85	0.06	1.00
First-Generation Immigrant	0.08	0.11	0.08	0.00	1.00
Second-Generation Immigrant	0.10	0.13	0.12	0.01	0.12
Wealth	0.19	0.26	0.19	0.00	0.95
AP Class	0.12	0.21	0.13	0.00	0.52
IB Program	0.02	0.02	0.02	0.77	1.00
Risk Factors	1.50	0.94	1.36	0.00	0.00
Parent Native English Speaker	0.85	0.81	0.83	0.00	0.17
School College Matriculation Rate	0.38	0.45	0.40	0.00	0.01
School Suspension Rate	0.21	0.12	0.16	0.00	0.00

Table 6.13: Balance Before and After Genetic Matching Within Schools (Matched Variables)

Variable	Mean Treatment	Mean Control		t-test p-values		KS-test p-values	
		Before	After	Before	After	Before	After
American Indian	0.01	0.01	0.01	0.47	0.25		
Asian	0.04	0.10	0.03	0.00	0.00		
Black	0.25	0.11	0.22	0.00	0.00		
Latino	0.19	0.14	0.19	0.00	0.93		
Multiracial	0.06	0.04	0.05	0.09	0.25		
Male	0.55	0.45	0.49	0.00	0.00		
Sports	0.44	0.54	0.49	0.00	0.02		
Math/Reading	-4.63	1.92	-3.91	0.00	0.00	0.00	0.04
Retain	0.24	0.10	0.17	0.00	0.00		
SES	-0.26	0.03	-0.24	0.00	0.51	0.00	0.51
Single Parent	0.34	0.21	0.25	0.00	0.00		
IEP: Yes	0.44	0.48	0.46	0.01	0.01		
IEP: Missing	0.44	0.47	0.47	0.20	0.00		
Parent College	0.26	0.39	0.22	0.00	0.00		
Urban	0.31	0.24	0.31	0.00	1.00		
Rural	0.22	0.24	0.22	0.36	1.00		
Northeast	0.14	0.17	0.14	0.02	1.00		
Midwest	0.21	0.25	0.21	0.01	1.00		
South	0.47	0.37	0.47	0.00	1.00		
School Math/Reading	-1.43	0.51	-1.43	0.00	1.00	0.00	1.00
School SES	-0.14	-0.03	-0.14	0.00	1.00	0.00	1.00

Table 6.14: Balance Before and After Genetic Matching Within Schools (Other Variables)

Variable	Mean Treatment	Mean Control		t-test p-values	
		Before	After	Before	After
Native English Speaker	0.85	0.83	0.82	0.06	0.03
First-Generation Immigrant	0.08	0.11	0.10	0.00	0.15
Second-Generation Immigrant	0.10	0.13	0.12	0.01	0.04
Wealth	0.19	0.26	0.23	0.00	0.03
AP Class	0.12	0.21	0.15	0.00	0.03
IB Program	0.02	0.02	0.02	0.77	0.69
Risk Factors	1.50	0.94	1.16	0.00	0.00
Parent Native English Speaker	0.85	0.81	0.83	0.00	0.10
School College Matriculation Rate	0.38	0.45	0.38	0.00	1.00
School Suspension Rate	0.21	0.12	0.21	0.00	1.00

Table 6.15: Descriptive Statistics for Coarsened Exact Matched Sample Across Schools (Matched Variables)

Variable	Proportion (%)	SE		
Suspension	14.34	0.47		
American Indian/Alaska Native	0.89	0.13		
Asian/Pacific Islander	1.97	0.18		
Black	11.06	0.42		
Hispanic	17.50	0.51		
More than one race	5.09	0.29		
White	63.48	0.64		
Male	48.61	0.67		
Participated in sports	53.11	0.67		
History of retention	9.30	0.39		
IEP status: Yes	3.45	0.24		
IEP status: No	49.77	0.67		
IEP status: Missing	46.78	0.67		
Single-parent household	19.89	0.53		
Parent college graduate	31.84	0.62		
Northeast	16.47	0.49		
Midwest	25.76	0.58		
South	38.60	0.65		
West	19.17	0.53		
Urban	23.03	0.56		
Suburban	51.85	0.67		
Rural	25.12	0.58		
	Mean	SD	Min	Max
Math and reading achievement	0.66	9.62	-28.50	27.40
Household socioeconomic status	-0.07	0.71	-1.97	1.98
School-level achievement	0.04	4.60	-14.20	17.35
School-level socioeconomic status	-0.07	0.35	-1.07	1.19

$(N = 5,622)$

Table 6.16: Descriptive Statistics for Coarsened Exact Matched Sample Within Similar Schools (Other Variables)

Variable	Proportion (%)	SE		
Suspension	35.91	1.97		
American Indian/Alaska Native	0.84	0.37		
Asian/Pacific Islander	0.34	0.24		
Black	9.23	1.19		
Hispanic	29.70	1.87		
More than one race	5.37	0.92		
White	54.53	2.04		
Male	50.34	2.05		
Participated in sports	46.98	2.05		
History of retention	7.21	1.06		
IEP status: Yes	1.68	0.53		
IEP status: No	54.70	2.04		
IEP status: Missing	43.62	2.03		
Single-parent household	12.25	1.34		
Parent college graduate	27.35	1.83		
Northeast	13.93	1.42		
Midwest	18.12	1.58		
South	43.62	2.03		
West	24.33	1.76		
Urban	25.67	1.79		
Suburban	58.22	2.02		
Rural	16.11	1.51		
	Mean	SD	Min	Max
Math and reading achievement	-1.61	9.89	-26.63	26.37
Household socioeconomic status	-0.17	0.82	-1.86	1.90
School-level achievement	-0.83	5.59	-14.20	16.06
School-level socioeconomic status	-0.12	0.44	-1.07	1.19

($N = 596$)

Table 6.17: Estimated SATT and Standard Errors

Matching Procedure	Across Schools (%)	Within Schools (%)
Propensity Score	-7.8 (1.7)	-12.4 (2.0)
Genetic	-11.3 (2.1)	-14.4 (2.1)
Coarsened Exact*	-10.2 (1.6)	-10.0 (3.4)
Mean	-9.8	- 13.4

Note. The coarsened exact matching estimates reflect a local SATT, as not all treated units were matched, and the within-school match was based on coarsened school characteristics rather than school identification number.

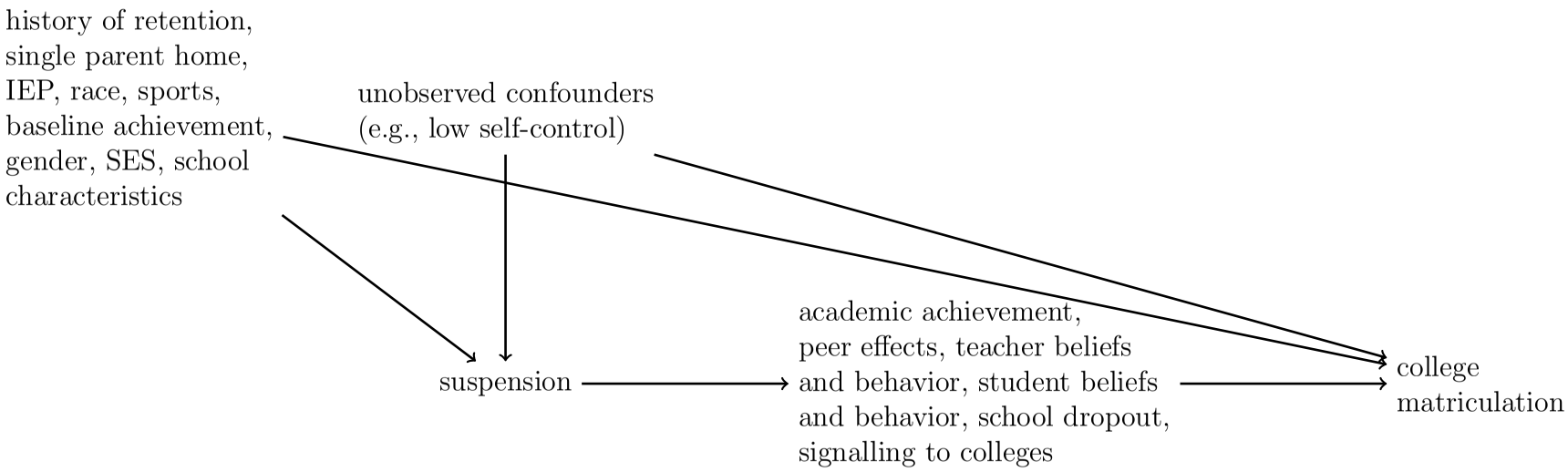

Figure 6.1: Conceptual Model

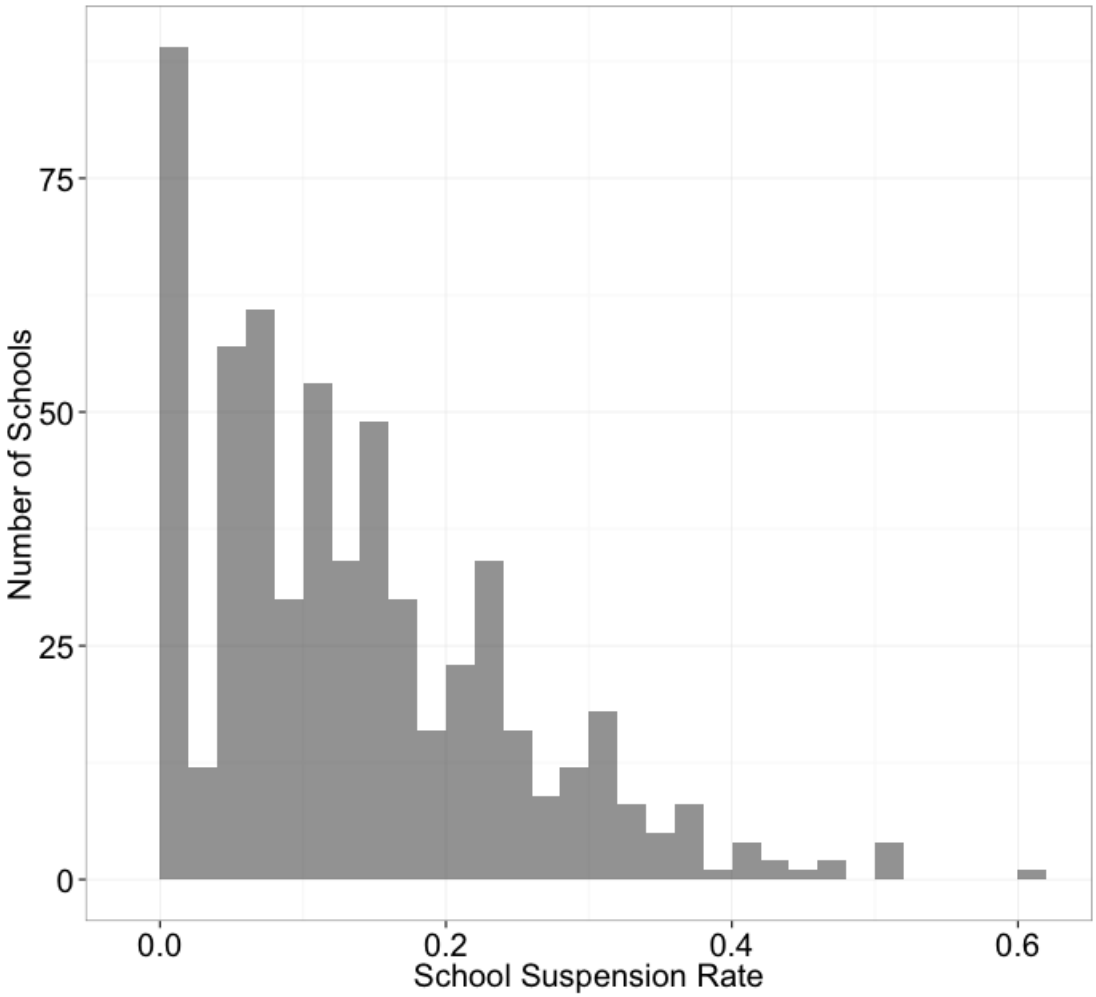

Figure 6.2: School Suspension Rate

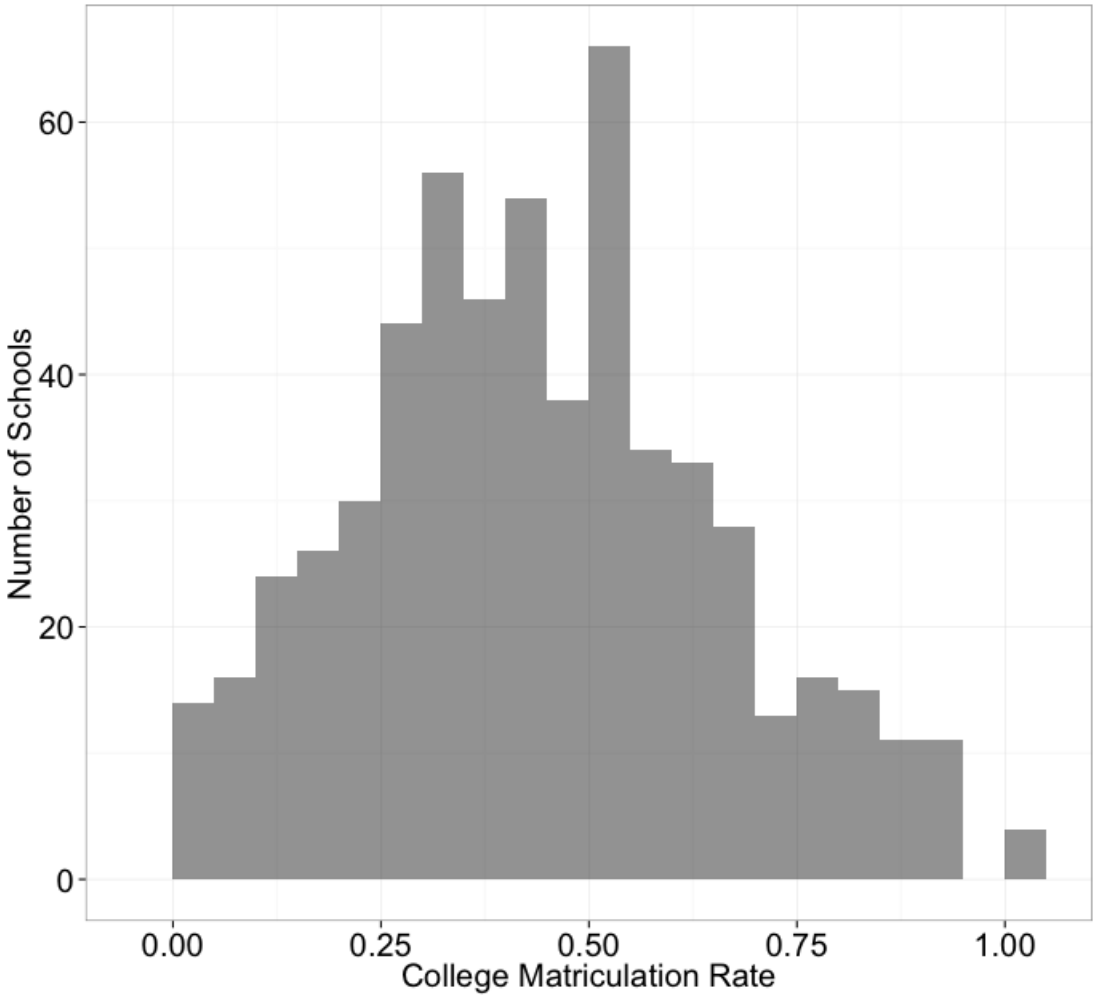

Figure 6.3: College Matriculation Rate

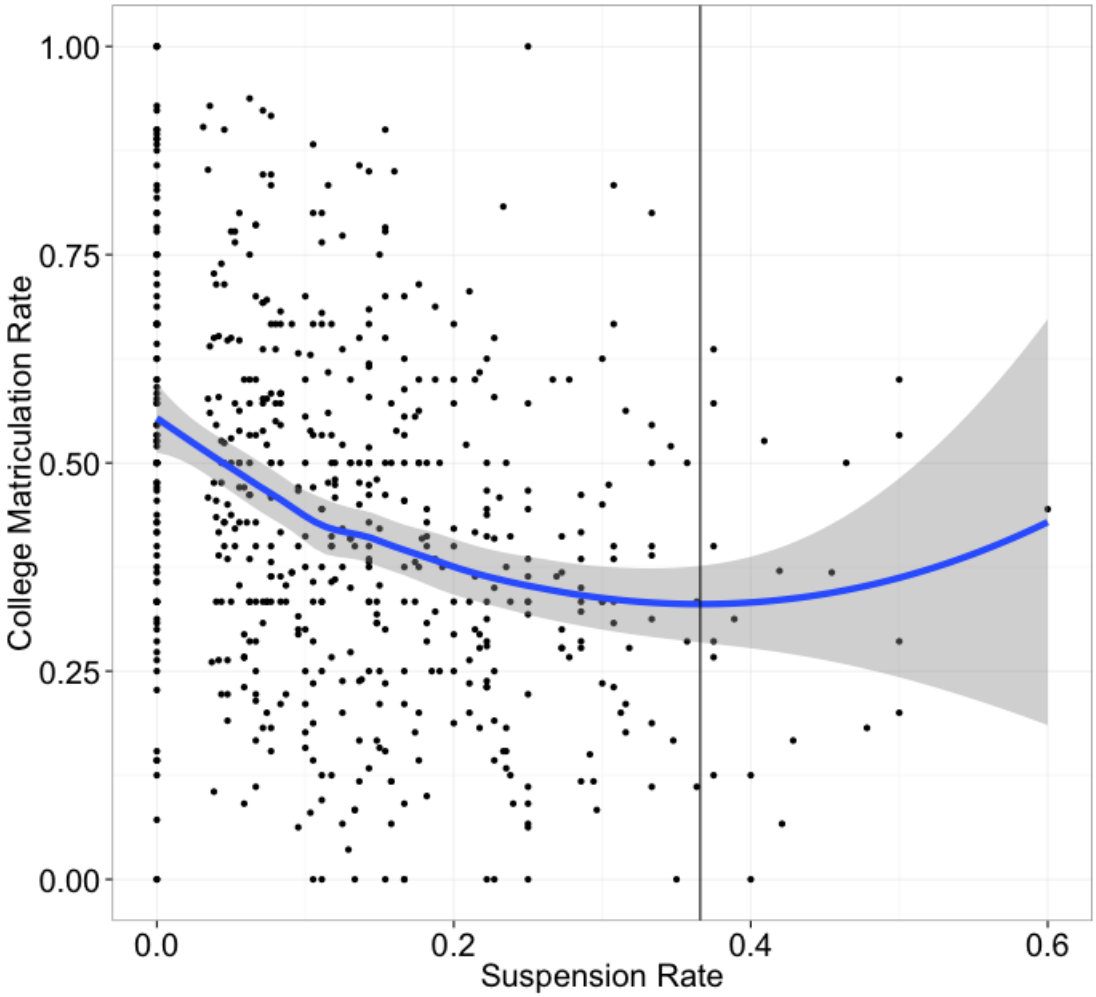

Figure 6.4: School-level Association Between Suspension and College Matriculation Rates

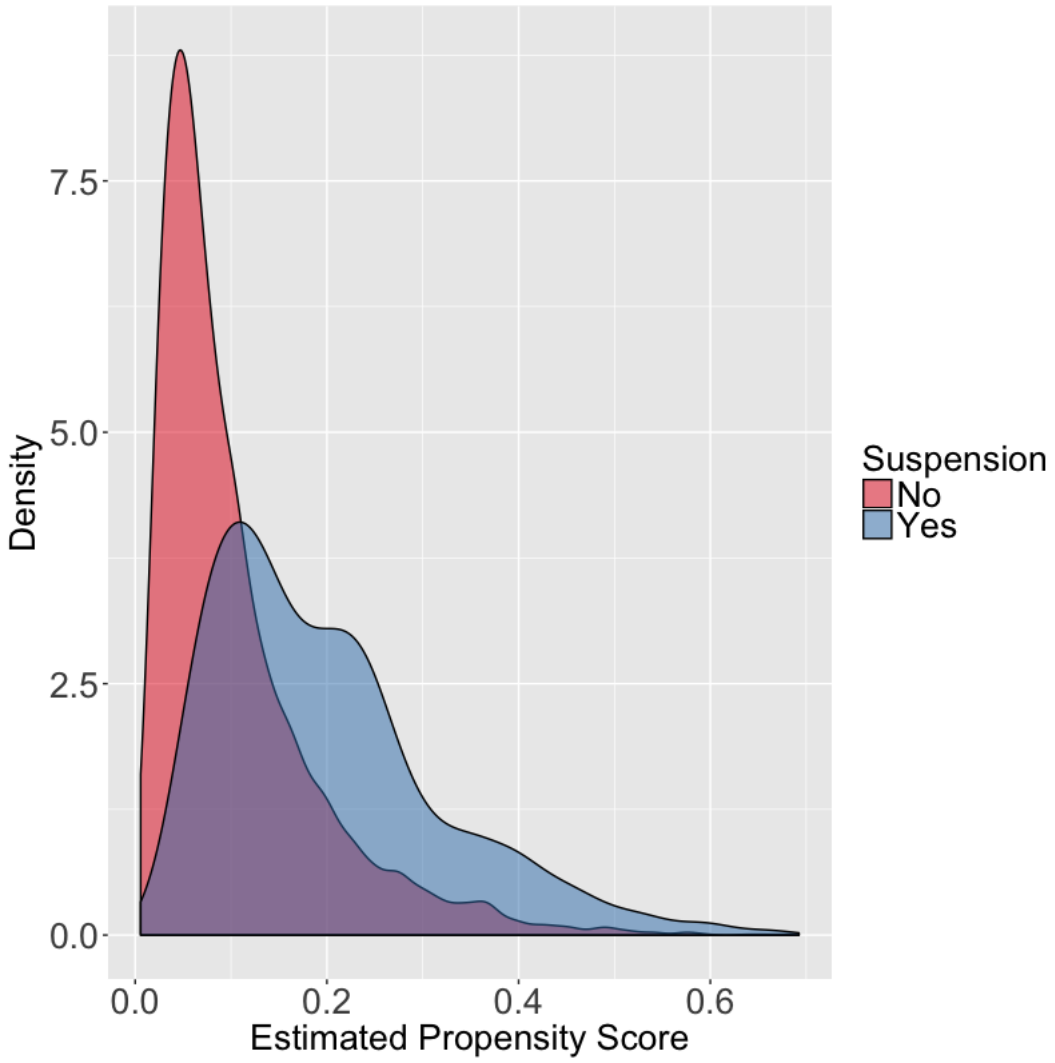

Figure 6.5: Common Support

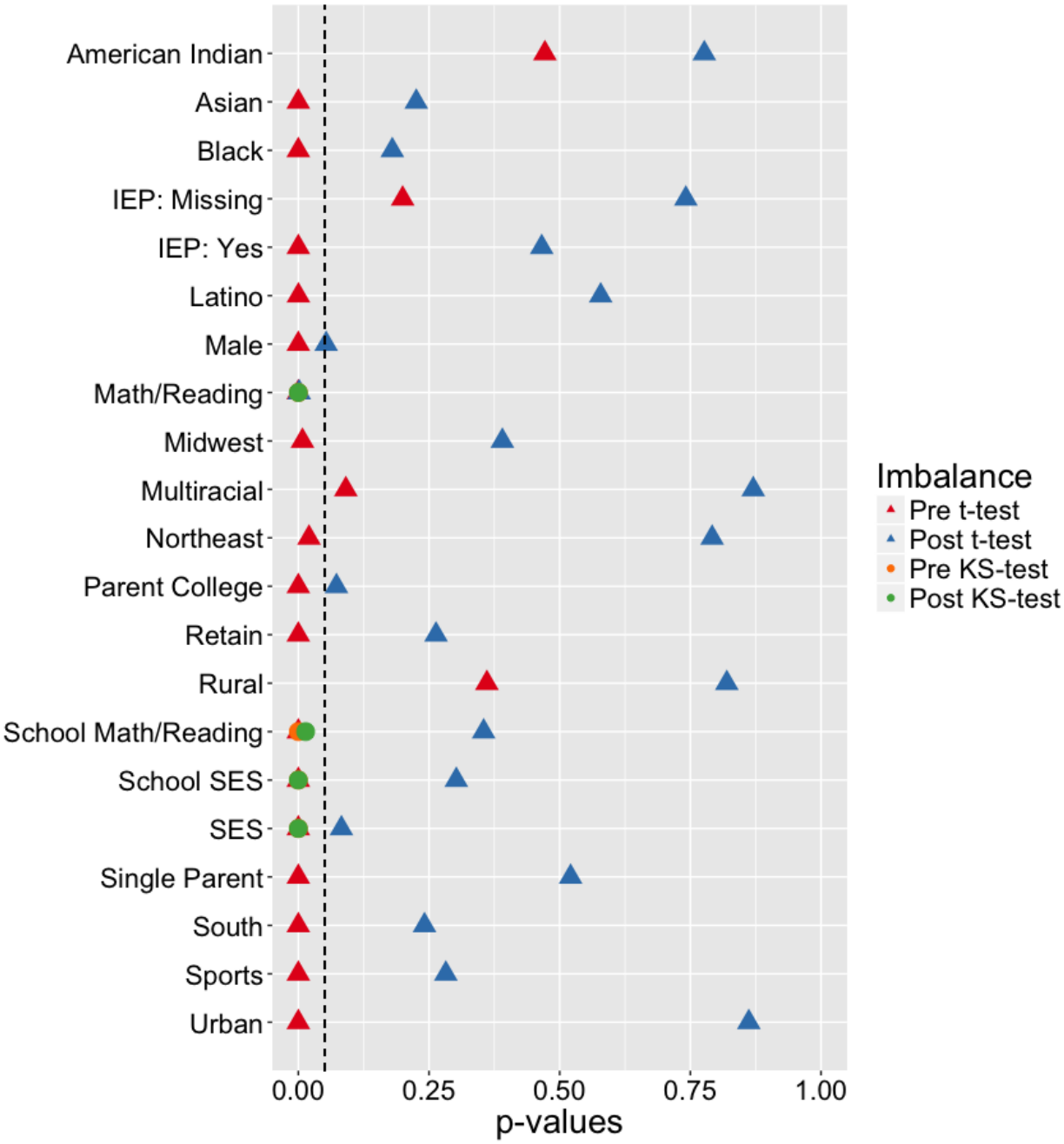

Figure 6.6: Balance After Propensity Score Matching Across Schools (Matched Variables)

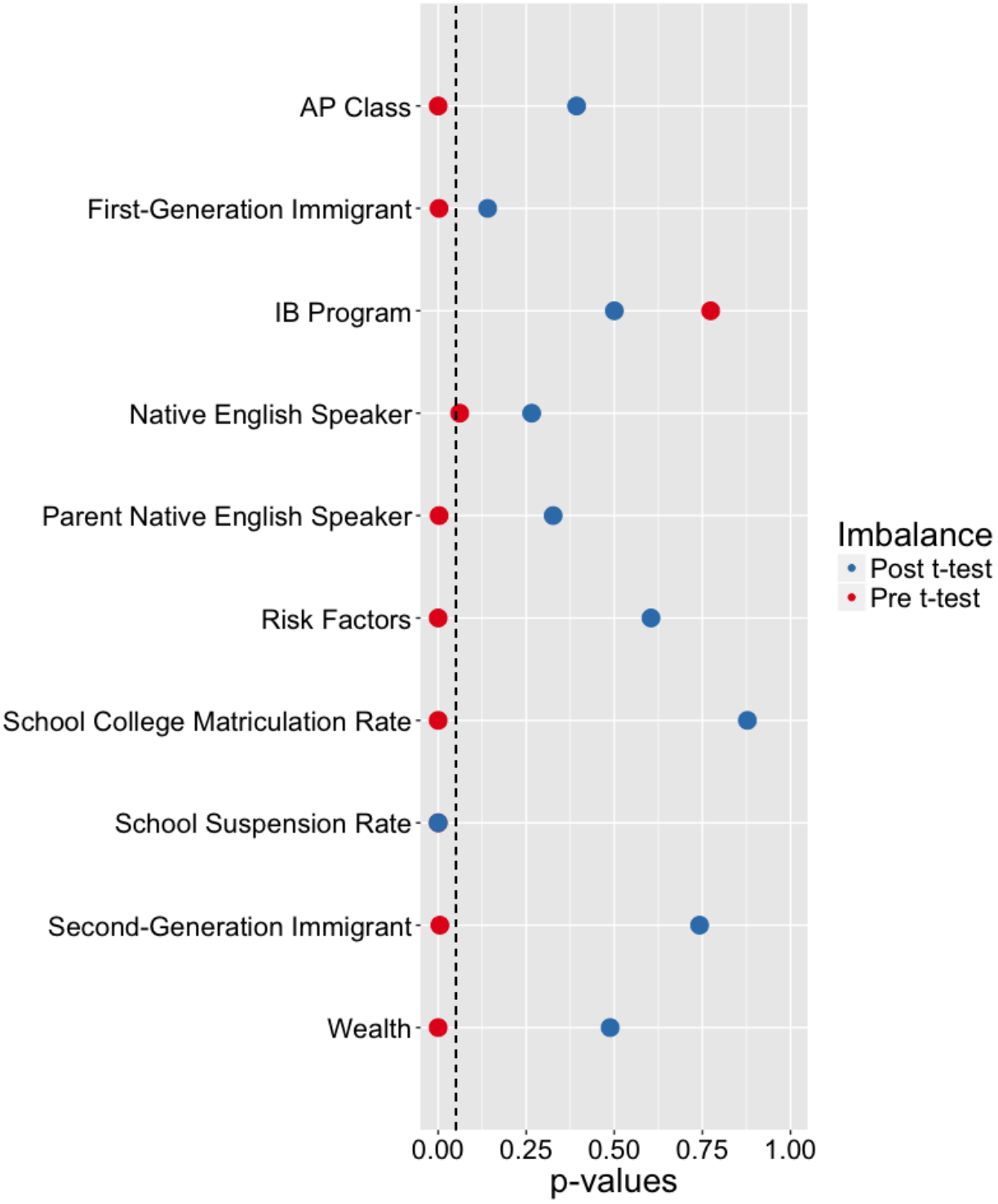

Figure 6.7: Balance After Propensity Score Matching Across Schools (Other Variables)

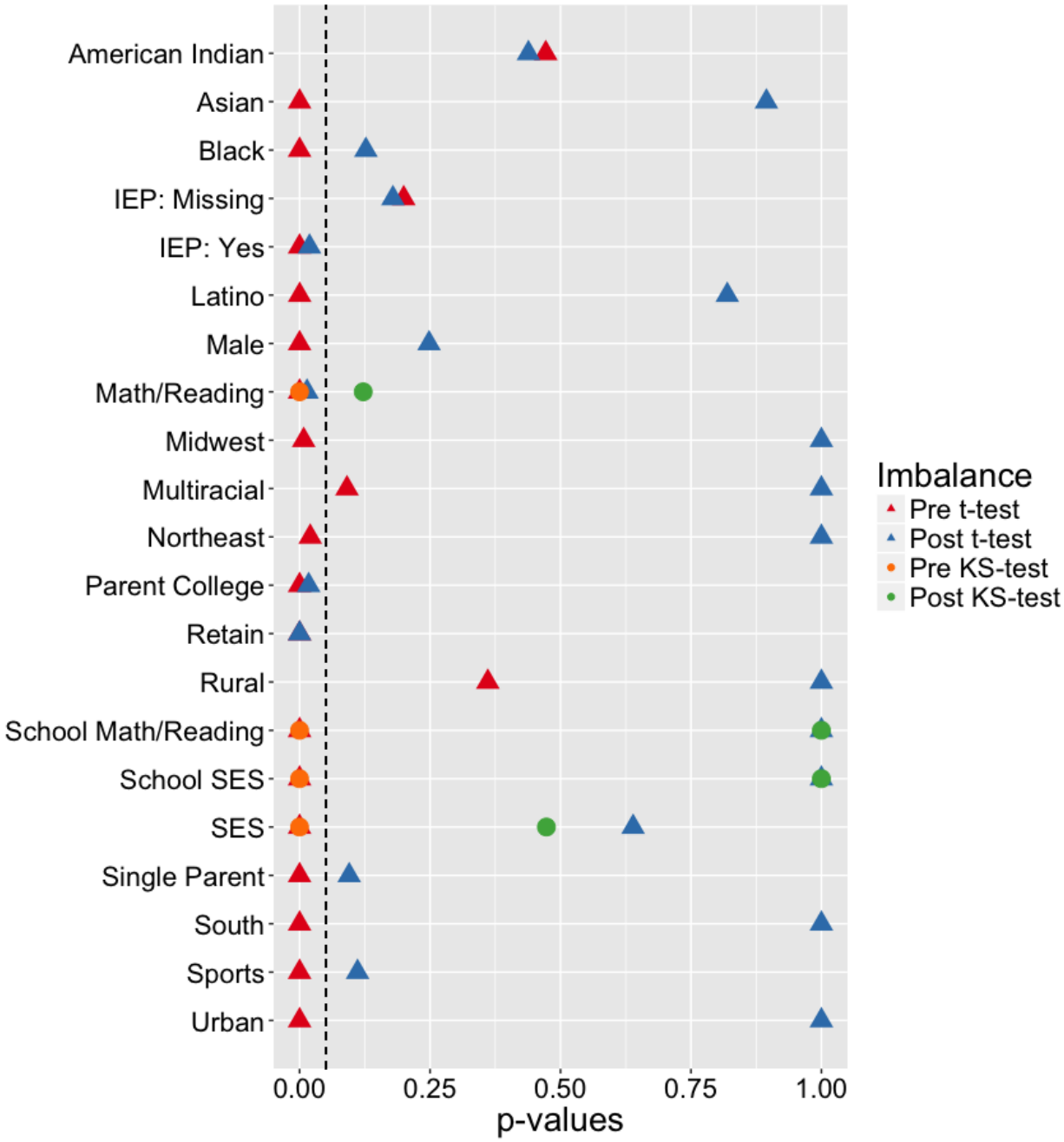

Figure 6.8: Balance After Propensity Score Matching Within Schools (Matched Variables)

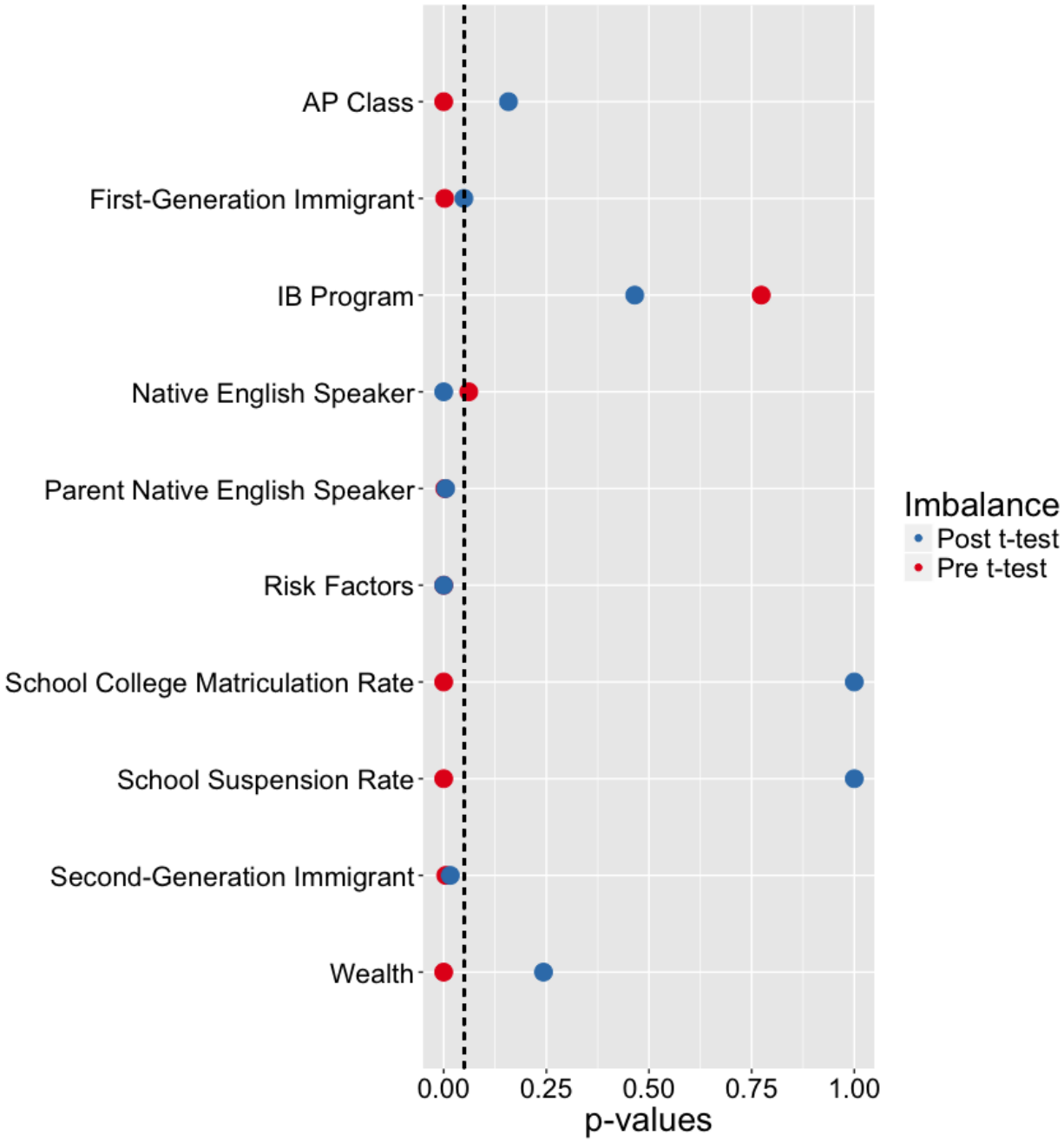

Figure 6.9: Balance After Propensity Score Matching Within Schools (Other Variables)

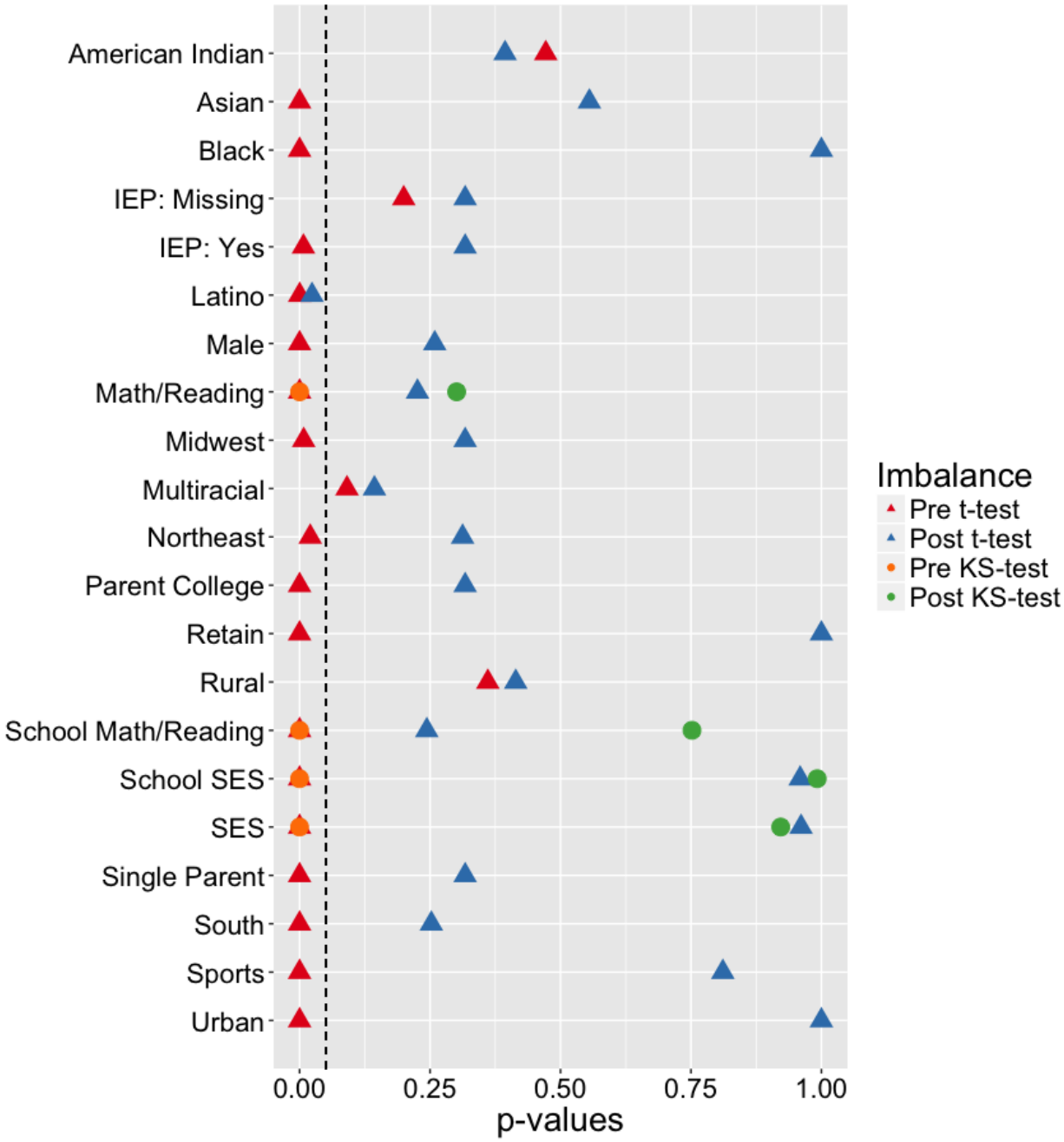

Figure 6.10: Balance After Genetic Matching Across Schools (Matched Variables)

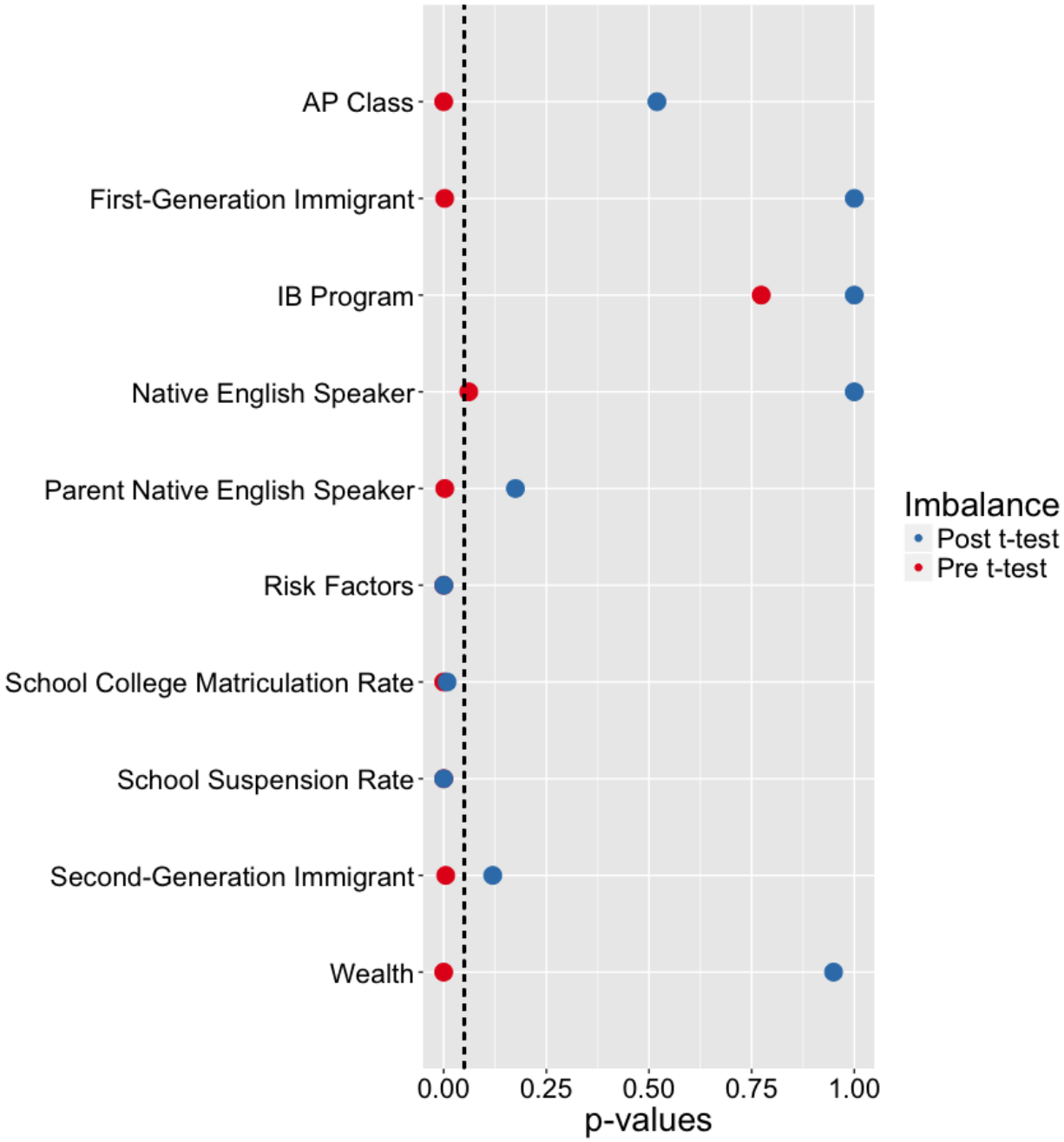

Figure 6.11: Balance After Genetic Matching Across Schools (Other Variables)

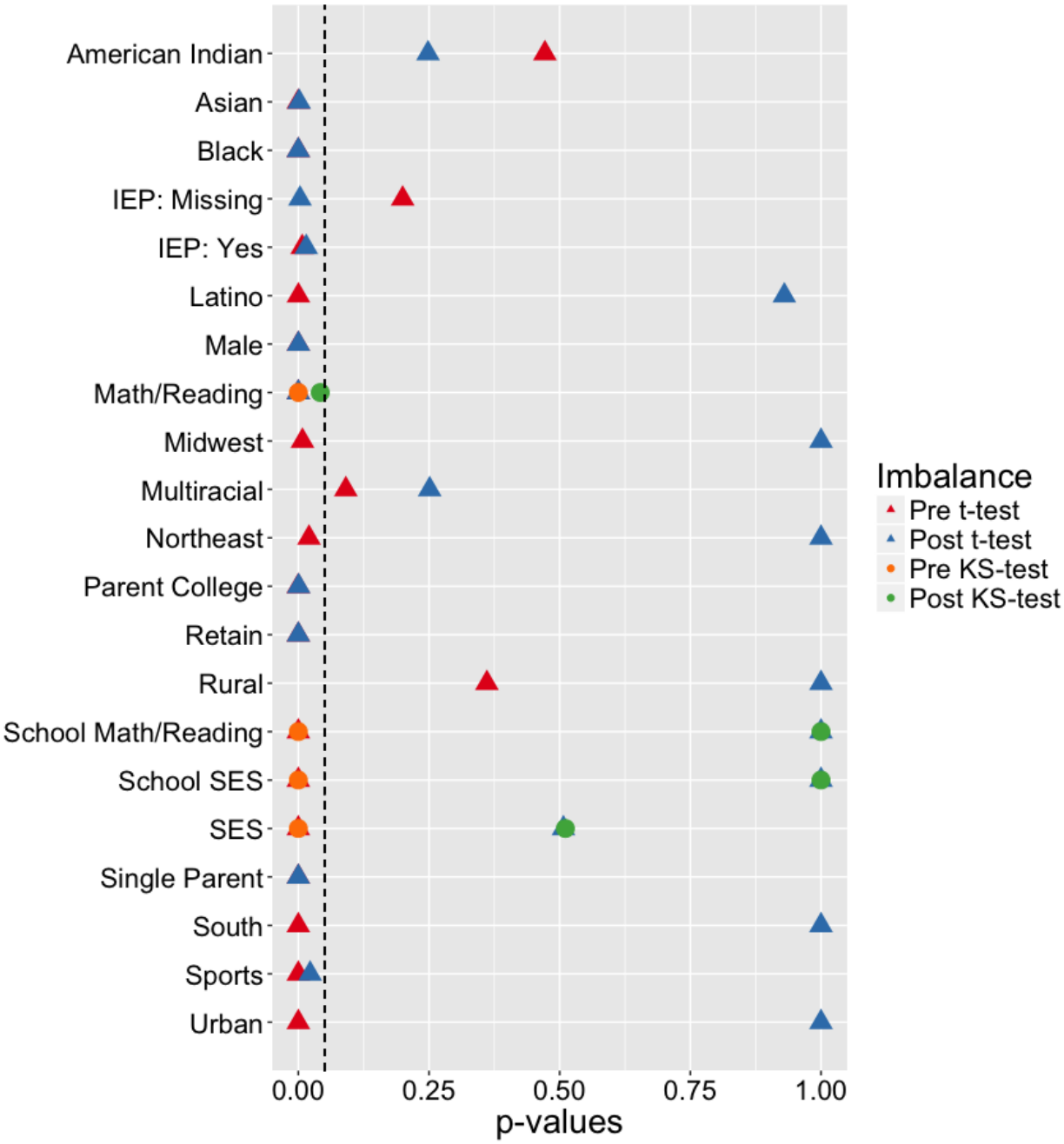

Figure 6.12: Balance After Genetic Matching Within Schools (Matched Variables)

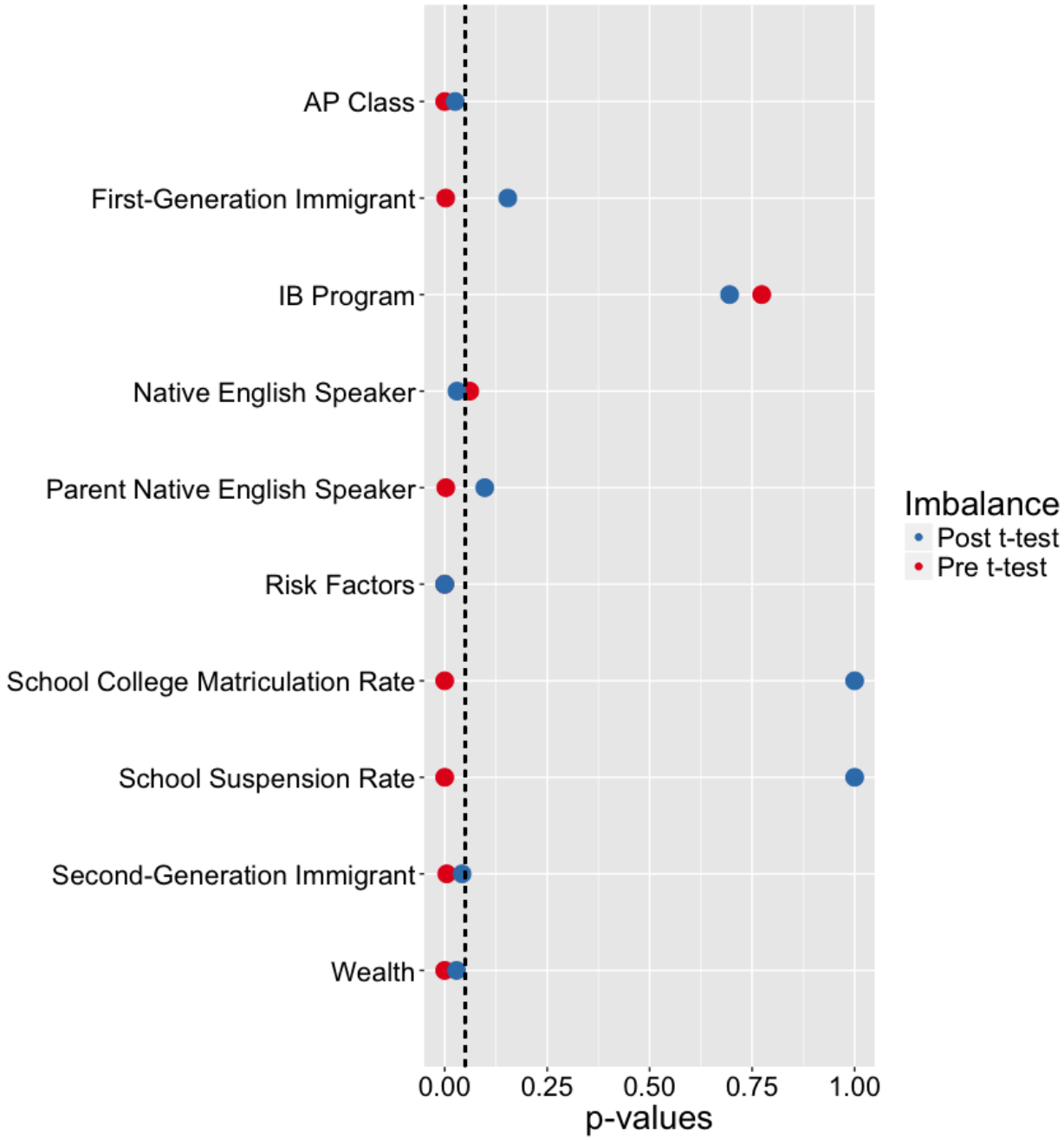

Figure 6.13: Balance After Genetic Matching Within Schools (Other Variables)