

NORTHWESTERN UNIVERSITY

A Better Understanding of School Suspensions: Bias, Policy Impact, and Student Experiences

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Human Development and Social Policy

By

Ijun Lai

EVANSTON, ILLINOIS

September 2018

© Copyright by (Ijun Lai) 2018

All Rights Reserved

Abstract

Recently, much public attention has been focused on racial inequalities in who is subjected to exclusionary school disciplinary policies, and consequently, forced to miss hours of instruction. Over the past two decades, researchers have documented the disparate impacts that zero-tolerance policies have had on students of color, low-income students, and students with disabilities (Advancement Project, 2000; Browne, 2003; Losen & Gillespie, 2012; Losen & Skiba, 2010; Skiba et al., 2002; Skiba & Peterson, 1999; Welch & Payne, 2010). Government officials at both the federal and local level are increasing efforts to guide school districts away from suspension policies and towards less punitive discipline strategies (Owens 2015, Willert 2015). National statistics show that there has been a decrease in out-of-school suspensions. During the 2015-2016 school year, 2.7 million students received one or more out-of-school suspensions--an almost 22 percent decrease from just four years earlier (US Department of Education, 2018). Yet, racial and gender disparities remain. For example, black male students make up 8 percent of the student population, but 25 percent of all students who received an out-of-school suspension. Comparatively, white females represent 24 percent of the student population, but account for only 8 percent of all students who received at least one out-of-school suspension (Ibid 2018). These racial disparities are particularly distressing since out-of-school suspensions are significantly correlated with poor attendance, lower academic achievement, high school dropout, juvenile justice system involvement, and unemployment (Advancement Project, 2000; Baker et al, 2001; Browne, 2003; Eaton, 2010; Fabelo et al, 2011; MacGillivray et al, 2008; Monahan et al., 2014; Sweeten, 2006; Wehlage & Rutter, 1986).

Despite growth in the literature on school suspensions and discipline, there are still several areas that are not well understood. My dissertation helps address some of these gaps. In my first study, I explore whether bias contributes to the current gender and racial disparities in out-of-school suspensions by examining how gender and race bias influences adult perceptions of youth as troublemakers. Additionally, I analyze how students' gender and race influence teachers' interpretations of and responses to students' actions. My second study builds a causal understanding of how Chicago's suspension reduction policy impacts student outcomes. The current literature has little empirical evidence of the causal relationship between suspensions and student outcomes. While previous studies link suspensions to bad academic and life outcomes, there is little to indicate that this relationship is causal. There could be many reasons why suspended students receive poor grades and have more interactions with the juvenile justice system that have nothing to do with suspension themselves. I use the difference-in-differences method to estimate the causal relationship between suspension reductions and freshmen year academic and nonacademic outcomes. Finally, my last study uses qualitative data to develop an understanding of students' suspension experiences and how they talk about fairness throughout the suspension process.

Acknowledgements

First and foremost, I am incredibly grateful for my dissertation chair and advisor, Dr. Jonathan Guryan, for guiding me through this journey. Thank you for all your support and wisdom throughout the years and showing me how excitingly impactful local research partnerships can be. Many thanks to my dissertation committee, Dr. Kirabo Jackson, Dr. Lincoln Quillian, and Simone Ispa-Landa, for always presenting new ways to think about my research and developing my skills as a scholar. A sincere thank you to Justina Schlund at Chicago Public Schools and the University of Chicago Crime and Education Labs, particularly Nathan Hess, for their support and assistance in accessing and using the necessary data for this dissertation. My deepest gratitude for the students and staff at Smith High School for opening their doors and a part of their lives to me.

I am also endlessly thankful for the amazing amount of support I have received from my classmates and family throughout the years. Thank you to Eric Brown, Katie Dahlke, Debbie Kim, Andrea Prado, Emily Ross, and Mary Clair Turner for feeding my stomach and spirit these past few years to make sure I can keep pushing forward. My deepest gratitude to my parents and my husband, Michael Tran, for taking wondrous care of me and my fur baby Lola. Finally, thank you to my remarkable support team—especially Mimi Do, Stephanie Kao, Jamie Kim, Ann Ko, Jane Ko, Justin Le, Jeffry Mutuc, and the Evanston Pogo Community—for bringing me laughter, helping me stay sane, and reminding me that there is life outside of graduate school.

Dedication

For all my students at Jordan Mott Junior High School and Bronx Writing Academy.

For all the teachers and staff working hard to create supportive school environments.

Contents

ABSTRACT.....	3
ACKNOWLEDGEMENTS	5
DEDICATION.....	6
CHAPTER 1: INTRODUCTION.....	13
Policy Background	15
Multi-leveled Approach	18
CHAPTER 2: GENDER, RACE, AND DISCIPLINE	20
Introduction.....	20
Theory	24
Empirical Strategy.....	26
Study 1: Troublemaker stereotype	28
Study 2: Teacher Perceptions.....	30
Gender Analysis	34
Racial Differences Amongst Males.....	36
Racial Differences Amongst Females	39
Robustness Checks.....	41

General Discussion and Conclusion	43
---	----

CHAPTER 3: SHORT-TERM IMPACTS OF CHICAGO’S SUSPENSIONS AND EXPULSIONS REDUCTION PLAN (SERP)..... 46

Introduction.....	46
-------------------	----

Suspension and Expulsion Reduction Plan	51
---	----

Theoretical Framework.....	53
----------------------------	----

Data	55
------------	----

Empirical Strategy.....	57
-------------------------	----

School Level Results	65
----------------------------	----

Student-Level Analysis.....	73
-----------------------------	----

High Risk Students	73
--------------------------	----

Lower Risk Students.....	76
--------------------------	----

Robustness Check	79
------------------------	----

Conclusion	83
------------------	----

CHAPTER 4: TALKING ABOUT DISCIPLINARY FAIRNESS 86

Introduction.....	86
-------------------	----

Data and Methods	87
------------------------	----

<i>School Context</i>	87
-----------------------------	----

	9
<i>Recruitment and sample</i>	<i>88</i>
<i>Interview Procedure</i>	<i>90</i>
<i>Coding for “fairness”</i>	<i>91</i>
<i>Social location/gaining rapport</i>	<i>91</i>
Findings	92
Discussion	107
TABLES.....	109
FIGURES.....	135
REFERENCES.....	151
APPENDIX TABLES	161
APPENDICES	171

List of Tables

TABLE 2.1. STUDY 1 SUMMARY STATISTICS	109
TABLE 2.2. TROUBLEMAKER STEREOTYPE GENDER DIFFERENCES.....	110
TABLE 2.3. TROUBLEMAKER STEREOTYPE WITHIN GENDER RACIAL DIFFERENCES	110
TABLE 2.4. STUDY 2 SUMMARY STATISTICS	111
TABLE 2.5. POOLED INCIDENT-SPECIFIC GENDER ANALYSIS	112
TABLE 2.6. INCIDENT-SPECIFIC GENDER ANALYSIS	113
TABLE 2.7. OVERALL IMPRESSION GENDER ANALYSIS	114
TABLE 2.8. RACIAL DIFFERENCES IN MALE INCIDENT-SPECIFIC OUTCOMES (POOLED).....	114
TABLE 2.9. RACIAL DIFFERENCES IN MALE INCIDENT-SPECIFIC OUTCOMES (BY INCIDENT).....	115
TABLE 2.10. RACIAL DIFFERENCES IN MALE OVERALL IMPRESSION	116
TABLE 2.11. RACIAL DIFFERENCES IN FEMALE INCIDENT-SPECIFIC OUTCOMES (POOLED)	116
TABLE 2.12. RACIAL DIFFERENCES IN FEMALE INCIDENT-SPECIFIC OUTCOMES (BY INCIDENT)	117
TABLE 2.13. RACIAL DIFFERENCES IN FEMALE OVERALL IMPRESSION.....	118

TABLE 3.1. OUT-OF-SCHOOL SUSPENSION PER STUDENT BY MODEL SPECIFICATION	119
TABLE 3.2. SCHOOL-LEVEL SUSPENSION OUTCOMES	120
TABLE 3.3. RACIAL BREAKDOWN OF OUT-OF-SCHOOL SUSPENSIONS	121
TABLE 3.4. SCHOOL AND ARREST OUTCOMES	123
TABLE 3.5. SUSPENSION OUTCOMES, HIGH RISK STUDENTS ONLY	124
TABLE 3.6. SCHOOL AND ARREST OUTCOMES, HIGH RISK STUDENTS ONLY	125
TABLE 3.7. SUSPENSION OUTCOMES, LOWER RISK STUDENTS ONLY	126
TABLE 3.8. SCHOOL AND ARREST OUTCOMES, LOWER RISK STUDENTS ONLY	127
TABLE 3.9. SCHOOL-LEVEL CHARACTERISTICS OVER TIME	128
TABLE 3.10. COMPOSITIONAL CHANGES IN "HIGH RISK" STUDENTS	130
TABLE 3.11. COMPOSITIONAL CHANGES IN "LOWER RISK" STUDENTS	132
TABLE 4.1. STUDY PARTICIPANT DETAILS	134

List of Figures

Figure 2.1. Incident Outcomes by Gender and Incident	135
Figure 2.2. Male Incident Outcomes by Race and Incident.....	136
Figure 2.3. Female Incident Outcomes by Race and Incident	137
Figure 3.1. OSS per Student by Year and Grade	138
Figure 3.2A. OSS per Student Treatment Difference over Time	139
Figure 3.2B. OSS Ever Treatment Differene over Time	140
Figure 3.2C. ISS per Student Treatment Difference over Time	141
Figure 3.2D. ISS Ever Treatment Differene over Time.....	142
Figure 3.3A. OSS Treatment Difference, High Risk Students Only	143
Figure 3.3B. OSS Ever Treatment Difference, High Risk Students Only.....	144
Figure 3.3C. ISS Treatment Difference, High Risk Students Only	145
Figure 3.3D. ISS Ever Treatment Difference, High Risk Students Only	146
Figure 3.4A. OSS Treatment Difference, Lower Risk Students Only.....	147
Figure 3.4B. OSS Ever Treatment Difference, Lower Risk Students Only	148
Figure 3.4C. ISS Treatment Difference, Lower Risk Students Only.....	149
Figure 3.4D. ISS Ever Treatment Difference, Lower Risk Students Only.....	150

Chapter 1: Introduction

Recently, much public attention has been focused on racial inequalities in who is subjected to exclusionary school disciplinary policies, and consequently, forced to miss hours of instruction. Over the past two decades, researchers have documented the disparate impacts that zero-tolerance policies have had on students of color, low-income students, and students with disabilities (Advancement Project, 2000; Browne, 2003; Losen & Gillespie, 2012; Losen & Skiba, 2010; Skiba et al., 2002; Skiba & Peterson, 1999; Welch & Payne, 2010). Government officials at both the federal and local level are increasing efforts to guide school districts away from suspension policies and towards less punitive discipline strategies (Owens 2015, Willert 2015b). National statistics show that there has been a decrease in out-of-school suspensions. During the 2015-2016 school year, 2.7 million students received one or more out-of-school suspensions--an almost 22 percent decrease from just four years earlier (US Department of Education, 2018)¹. Yet, racial and gender disparities remain. For example, black male students make up 8 percent of the student population, but 25 percent of all students who received an out-of-school suspension. Comparatively, white females represent 24 percent of the student population, but account for only 8 percent of all students who received at least one out-of-school suspension (Ibid 2018). These racial disparities are particularly distressing since out-of-school suspensions are significantly correlated with poor attendance, lower academic achievement, high school dropout, juvenile justice system involvement, and unemployment (Advancement Project,

¹ 3.45 million students received at least one out of school suspension during the 2011-2012 school year (US Department of Education, 2014)

2000; Baker et al, 2001; Browne, 2003; Eaton, 2010; Fabelo et al, 2011; MacGillivray et al, 2008; Monahan et al., 2014; Sweeten, 2006; Wehlage & Rutter, 1986).

Despite growth in the literature on school suspensions and discipline, there are still several areas that are not well understood. My dissertation helps address some of these gaps. In Chapter 2, I explore whether bias contributes to the current gender and racial disparities in out-of-school suspensions by examining how gender and race bias influences adult perceptions of youth as troublemakers. Additionally, I analyze how students' gender and race influence teachers' interpretations of and responses to students' actions. Next, Chapter 3 builds a causal understanding of how Chicago's suspension reduction policy impacts student outcomes. The current literature has little empirical evidence of the causal relationship between suspensions and student outcomes. While previous studies link suspensions to bad academic and life outcomes, there is little to indicate that this relationship is causal. There could be many reasons why suspended students receive poor grades and have more interactions with the juvenile justice system that have nothing to do with suspension themselves. Chapter 3 utilizes the difference-in-differences method to estimate the causal relationship between suspension reductions and freshmen year academic and nonacademic outcomes. Finally, Chapter 4 uses qualitative data to develop an understanding of students' suspension experiences and how they talk about fairness throughout the suspension process.

In the remainder of the introduction, I briefly review the history of zero tolerance policies and what the current literature says about different factors involved in the suspension process.

Policy Background

As television coverage of juvenile homicides and school violence increased from the late 1980s to the early 1990s, parents blamed “soft” rehabilitative approaches and called for harsher punishments to increase school security (Moore et al, 2003). At the same time, the “broken windows” philosophy was quickly gaining public support. This theory asserted that immediate and severe consequences for all crimes, even minor crimes like breaking windows, are necessary to prevent future crimes (Wilson & Kelling, 1982). School districts rapidly adapted this mentality in new policies (Brady et al, 2007; Drum Major Institute, 2005). To address these concerns on a national level, President Clinton signed into law the Improving America’s Schools Act of 1994 (Public Law 103-382). This legislation included the Gun-Free Schools Act, which required schools to have zero tolerance for firearms in schools by expelling students for at least one year if they bring a firearm to school and to refer the student to the local criminal or juvenile justice system. Non-compliant schools lose all federal funding. Importantly, the law requires schools districts to have a chief administering officer that can review expulsions on a case-by-case basis. Rather than interpreting this law as being flexible towards situational contexts, many schools took this law as an opportunity to expand harsh punishments for a wide variety of actions including cursing, disobedience (Browne, 2005; Giroux, 2009).

In 1997, the language of the Gun-Free Schools Act expanded from firearm to the more generic term “weapons” and 79% of public schools had zero tolerance policies in place to deal with fights (Casella, 2003). Since the change, schools have expanded the definition of weapons to include: squirt gun, wallets attached to chains, and an egg (Browne, 2003; Skiba & Knesting, 2001). The combination of increased police and school resource officers in schools and

increasingly vague zero tolerance policies was rapidly followed by a dramatic rise in suspensions, expulsions, school-based arrests, and juvenile court appearances across the country. Ethnographic studies noted how “fighting in the hallway is classified as assault; swiping a classmate’s pencil case can be classified as a property crime; and talking back to a school security officers or being late to a class is disorderly conduct” (Mukherjee, 2007 p. 18). There were over 8,500 Chicago Public Schools school-based arrests in 2003. In over 40% of the arrests, the only weapons cited were hands or feet used in minor fights with no major injuries (Browne, 2005; Advancement Project, 2000). This pattern was repeated across the country with the majority of arrests due to simple assaults (fights) or miscellaneous (usually charges like disturbing the peace) incidents that would have been resolved with parent-teacher conferences or afterschool detentions in the past but are now quickly escalated because of the prevalence of officers and zero tolerance policies. Drug violations and weapons possession usually made up only 5-20% of arrests (Browne, 2003; Skiba & Rausch, 2006).

While the impetus for zero tolerance centered around school safety, researchers quickly found that most school suspensions were due to classroom disruptions (Bracy, 2011; Fenning & Rose, 2007; Vavrus & Cole, 2002). The majority of suspensions involved multiple challenges to the teacher’s ability to maintain control of the classroom. Teacher surveys show that teachers who don’t trust their students are more likely to rely on harsh classroom management techniques (Eccles, 1999; Eccles et al, 1993). Classroom disruptions are commonly seen as a challenge to a teacher’s authority. Suspensions and expulsions are a very public way for teachers to reassert their power. Qualitative studies find that to regain control of the class, teachers would single out one student to be disciplined. When the same exact student engaged in other low-level

disruptions when the teacher felt in control of the class, these incidences were often completely ignored (Fenning & Rose, 2007; Vavrus & Cole, 2002).

Given the growing body of literature documenting correlations between suspensions and bad outcomes such as lower grades and higher dropout rates, as well as reports that students, particularly black students, were being suspended for minor infractions, public pressure against suspension usage began to grow. Government officials at both the federal and local level are increasing their efforts to help school districts turn away from exclusionary policies and towards alternative discipline (Owens 2015, Willert 2015b). In 2011, the US Departments of Justice and Education created the national Supportive School Discipline Initiative to spearhead discipline-related policy coordination, data collection and research efforts, and provide guidance on strategies such as restorative justice and positive behavior interventions (US Department of Justice/Department of Education, 2011). As part of the Every Student Succeeds Act (passed in 2015 and replacing the No Child Left Behind Act), states are required to track suspension rates and create plans to decrease their suspension use (Goldstein, 2017). Currently, 27 states and more than 50 of America's largest school districts have created significant changes to their student codes of conduct and discipline policies to include nonpunitive discipline strategies and/or limits suspension usage (Eden 2017; Steinberg and Lacoë 2017). For example, in 2014, California became the first state to ban suspensions for minor misbehaviors such as "willful defiance" (California, Education Code § 48900(k))—a term often used for infractions such as refusing to take off a hat or not giving up a cell phone (Watanabe, 2013).

Multi-leveled Approach

Studies have shown that suspensions are context-sensitive. A child's risk of suspension relies on the interaction between school beliefs, teacher beliefs, and student characteristics (Morrison et al, 2001). Within the same school district, there is large variation in how many suspensions occur in each school. Principal surveys show that the number of suspensions per school largely depend on each principal's school discipline philosophy (Advancement Project, 2000; Dunbar & Villarruel, 2002; Fabelo et al, 2011; Skiba et al, 2014). Teachers at schools where the principal emphasizes prevention and intervention know that the principal will deny their requests for suspension. Consequently, they are less likely to send their students to the office and there are few suspensions at these schools (Skiba et al, 2003). Teachers have a lot of latitude in deciding how to handle classroom disruptions—whether incidences should be handled in the classroom or not. At the average school, only a handful of teachers are responsible for most office referrals (Skiba & Rausch, 2006). Finally, punishments vary by students' own academic and behavioral histories. Students who are labeled low-achievers or troublemakers are more likely to face harsh punishment (Ferguson, 2001; Morrison & D'Incau, 1997). Multi-level analysis of disciplinary incidences found that all three sets of characteristics (school, teacher, and student) accounted for more variation in the probability of being suspended than details of the actual incident (Vavrus & Cole, 2002; Wu et al, 1982).

My dissertation takes a closer look at the suspension process from these three different levels. The second chapter of my dissertation focuses on how teachers' perceptions of students' actions may vary based on students' gender and race. This study provides insight into one possible mechanism driving the gender and racial disparities in out-of-school suspensions.

Chapter 3 evaluates the impact of a citywide suspension reduction policy from two perspectives—schoolwide averages and student-level risk groups. The policy created more intense focus on high suspending schools, suggesting great shifts in school approaches to suspensions. At the same time, it is possible that this policy had a differential impact on students depending on their individual suspension-risk levels. For example, students with a history of suspension or poor attendance may experience the policy change in a divergent way than students who do not have a previous suspension history and/or has great attendance and grades. The second half of Chapter 3 delves into this heterogeneity. Lastly, the final chapter of my dissertation documents how students talk about fairness throughout the suspension process. The data draws from a larger qualitative project designed to build comprehension of students' suspension experiences. These experiences can elucidate the mechanisms behind why suspensions do or do not have impacts on students' academic and non-academic outcomes. Perceptions of fairness is one aspect that may influence how students perceive of their suspension experiences.

Chapter 2: Gender, Race, and Discipline

Introduction

According to the most recent data from the US Department of Education Office for Civil Rights, black boys are significantly overrepresented in suspension and expulsion statistics (2016). 18% of all black boys received out-of-school suspensions during the 2013-2014 school year, compared to 5% of white boys. With much of the attention focused on boys, racial disparities among girls are often overlooked. 10% of black girls received out-of-school suspensions—five times the rate of white girls (2%) and *double* the rate of white boys (Ibid, 2016). These disparities are particularly worrisome since suspensions are significantly correlated with lower academic achievement, high school dropout, juvenile justice system involvement, and unemployment (Advancement Project, 2000; Baker et al, 2001; Browne, 2003; Eaton, 2010; Fabelo et al, 2011; MacGillivray et al, 2008; Monahan et al., 2014; Sweeten, 2006; Wehlage & Rutter, 1986).

Previous researchers have uncovered potential contributors to the disciplinary racial gaps. One group of researchers argue that disciplinary racial differences are simply due to black students' higher misbehavior rates compared to their white counterparts (Beaver et al, 2011; Huang, 2016; Huang, 2017; Rocque, 2010; Wright et al., 2014). Others argue that black students misbehave at the same rates as white students, but are subjected to stricter surveillance and increased likelihood of punishment (Ferguson, 2000; Gilliam et al., 2016; Morris, 2005; Morris, 2007; Okonofua & Eberhardt, 2015; Rios, 2011; Wun, 2016). Yet, few studies have focused on the unique role gender, both overall and in combination with racial groups, may play in

stereotypes of youth as troublemakers, perceptions of misbehaviors, as well as disciplinary decisions.

Our study helps fill this current gap in the literature by using a series of survey experiments to measure gender differences as well as racial differences within gender of: 1) preconceived notions of youth as troublemakers and 2) how minor misbehaviors are perceived by teachers and inform their disciplinary decisions. Participants are randomly assigned to black/white and male/female treatment conditions. The first study asked participants to estimate the average Americans' view of youth (black girl, black boy, white girl or white boy) across a series of characteristics—politeness, troublemaking, sassy, and angry. The second study provided teachers with a vignette of a minor classroom disturbance by a black/white male/female student and asked teachers to evaluate students' actions (in terms of severity, hindrance, irritation, and discipline). Teachers are then given a second minor classroom disturbance and asked to re-evaluate students' actions before providing their overall impression of the student. We use these experimental designs to answer the following research questions:

Study 1

- 1) Are there gender differences in preconceived notions of youth as troublemakers?
- 2) Within each gender group, are there racial differences in preconceived notions of youth as troublemakers?
 - a. Are there racial differences amongst boys?
 - b. Are there racial differences amongst girls?

Study 2

- 3) Are there gender differences in teachers' perceptions of and responses to students' minor misbehaviors?
- 4) Within each gender group, are there racial differences in teachers' perceptions of responses to students' minor misbehaviors?
 - a. Are there racial differences amongst boys?
 - b. Are there racial differences amongst girls?

We hypothesize that females, particularly white females, are generally viewed as less troublesome than their male counterparts. Additionally, we hypothesize that some of this goodwill towards girls, particularly white girls, may spill over into how teachers perceive students' minor misbehaviors. Specifically, we theorize that teachers will be less bothered by female students' actions.

Consistent with our theory, we find that adults' views of youth as troublemakers vary significantly by gender and race. Males were 0.215 standard deviations more likely to be viewed as troublemakers in comparison to their female counterparts. Within males, black males were 0.732 standard deviations more likely to be viewed as troublemakers in comparison to their white male peers. Similarly, black females were viewed as 0.711 standard deviations more troublemaking than white females.

These general stereotypes may play a role in school discipline. We find that, for the exact same first misbehaviors, males' actions were viewed as more hindering (0.283 standard deviations) and more irritating (0.258 standard deviations) than female offenders. Correspondingly, teachers advocated for harsher consequences for males (0.236 standard deviations with 0.061 p-value), though this estimate was not statistically significant at the 0.05 alpha-level. Notably, these gender differences disappeared following the second disciplinary incident. After reading both disciplinary incidents, teachers were asked to evaluate their overall impression of students across five characteristics—impolite, troublemaker, angry, problematic pattern of behavior, and dangerous. We find no evidence of significant gender differences across these characteristics, though point estimates suggest teachers were more likely to label male students as troublemakers, angry, and dangerous.

Within male students, we find evidence of racial bias against black male students. Teachers reported feeling significantly more hindered (0.404 standard deviations) and irritated (0.583 standard deviations) by black male students' actions after the first disciplinary incident. Following the second disciplinary incident, teachers reported no significant racial differences in feeling hindered, but they continued to feel more irritated by black males compared to their white male peers (0.440 standard deviations). We find no significant differences, after either disciplinary incident, in how severely students should be disciplined. Our estimates suggest that teachers were more likely to assign negative characteristics to black male students than their white male counterparts. Teachers were significantly more likely to label black male students as impolite (0.419 standard deviations), a troublemaker (0.432 standard deviations), and to believe that their actions were indicative of a pattern (0.424 standard deviations). They were not significantly more likely to see black male students as angry nor dangerous.

Contrary to both our hypothesis and our findings from Study 1, we find no significant racial differences in how teachers perceived female misbehaviors. While not significant at the traditional 0.05 p-value, our point estimate suggests a bias against black females about how severely students should be disciplined following the first misbehavior (0.367 standard deviation point estimate, with a 0.07 p-value). There does not appear to be any significant racial differences in discipline recommendation following the second disciplinary infraction. Furthermore, there are no significant racial differences in teachers' overall impression of female students.

The remainder of the paper is as follows. Section 2 presents our theoretical framework and Section 3 describes our experimental design. Section 4 details the design, sample, and

analysis for Study 1 while Section 5 repeats the process for Study 2. Section 6 provides a robustness check and Section 7 concludes.

Theory

A. Gender and Discipline

According to national statistics and numerous suspension studies, white females are significantly less likely to receive suspensions and expulsions (Blake et al., 2011; Mendez et al., 2002; US DOE, 2016). One explanation for this difference may lie in benevolent sexism (Glick & Fiske, 1996). Female students may be benefiting from school officials' implicit belief that girls are good and treat their behavioral transgressions as trivial matters that occur simply because girls are overly emotional and dramatic (Greenwald & Banaji, 1995; Waldron, 2011). Indeed, studies often find that misbehaviors among girls, such as name-calling, gossiping, social exclusion, are often ignored or downplayed by many adults. Teachers, other adults, and adolescent girls themselves may explain away such behaviors by saying that these actions are just part of a normal developmental phase (Merten, 1997; Simmons, 2002). In Waldron's 2011 ethnographic study of two high schools, she found a stark gender differential in how students were punished for fights. Girl fights were often resolved through referrals to peer mediation while fights amongst boys often resulted in detentions and suspensions. It is unclear from her study whether this difference was due to variation in the physicality of the fights or if it can simply be explained by participants' gender.

Alternative theories suggest that given the same infraction, female students may receive harsher punishments. Supporters of this theory argue that schools use discipline to socialize youth into traditional gendered and racial roles (Giroux & Purpel, 1983; Martin, 1998; Raby,

2005). The traditional female stereotypes are of someone who is accommodating, speaks softly, and is subordinate (Eberhardt & Fiske, 1994; Fiske 2010; Glick & Fiske 2001). When girls do not behave according to these gender roles, they may be quickly punished. This process can begin as early as pre-school. Martin (1998) observed five preschool classrooms and found that teachers monitored and disciplined girls' appearance and behaviors while boys were given much more latitude in their actions. For example, boys were often running around and crashing into people but teachers would only intervene when there was an immediate danger. On the other hand, if girls spoke or argued loudly, they were immediately reprimanded and told to be softer and quieter. If similar processes are occurring throughout all grades, we hypothesize that females may be punished, and in some cases, receive harsher punishments, for minor disruptions actions (such as throwing things and making loud noises) because they are perceived as less ladylike.

B. Intersectionality: Black Females and Discipline

We also hypothesize that perceptions of students' behaviors, and subsequent disciplinary decisions, may vary between white girls and black girls. In a recent analysis of suspension and expulsion rates in New York City, researchers found that while black boys were five times more likely than white boys to be suspended, black girls were *ten* times more likely to be suspended than their white counterparts (Crenshaw et al., 2015). Intersectionality argues that multiple group identities create unique experiences, filled with both advantages and disadvantages (Purdie-Vaughns & Eibach, 2008). Our experiment offers a unique lens to see how gender and race may interact in school discipline decisions.

Black girls' experiences at school are complex and contradictory. On the one hand, teachers expect black females to outperform their black male counterparts (Morris, 2005; Tenenbaum & Ruck, 2007). Part of succeeding in school means actively participating in classes,

including answering and asking questions. Yet when black girls speak up in class, their questions are often seen as disrespect and/or they are quickly labeled as loud, having an attitude, too assertive and impolite (Koonce, 2012; Morris, 2005; Morris, 2007; Morris, 2016; Waldron, 2011). Furthermore, ethnographic data document how teachers often ignore black female students when they shout out answers while acknowledging black male students' answers when they do the same. These teacher reactions can lead some students to disengage from class or act out in frustration (Evans, 1988; Fordham, 1993; Morris, 2007; Wun, 2016). It seems that if they want to succeed, black female students need to actively subscribe to the female stereotype of being silent and obedient (Fordham, 1993; Morris, 2007).

Similar to the literature on black boys, qualitative studies have consistently found that black girls' movements are under strict surveillance at school. Students often note that they are quickly called out and punished for minor behaviors, such as having a beverage in class or using lip balm, that are ignored when done by their white peers (Morris, 2005; Wun, 2016). However, the black female experience also varies from the black male experience in gender-specific ways. Qualitative work suggests that black girls are particularly singled out for transgressions of gender norms. Observation notes and interviews revealed that black girls were constantly scolded for not sitting properly, running, told how to dress properly (clothes cannot be too tight or too short), and told to speak quietly (Morris 2005, 2007; Wun, 2015).

Empirical Strategy

Although our two studies ask different questions and use two different samples, the research design is identical. Respondents are randomized into one of four different treatment

conditions: black boy, white boy, black girl, and white girl. We begin by measuring gender differences with the following equation:

$$(1) \quad Y_i = \beta_0 + \beta_1 \text{Male}_i + X_i \beta_2 + e_i$$

where Y is the outcome of interest and i indexes individual respondents. Male_i is an indicator variable equal to 1 if the respondent was randomly assigned to the male treatment condition. Thus, β_1 represents the male-female student difference in the outcome of interest. To improve precision, we include X , a vector of participant demographics².

Next, we estimate the racial differences within each gender group using the following equations:

$$(2a) \quad Y_i = \beta_0 + \beta_1 \text{BlackMale}_i + X_i \beta_2 + e_i$$

$$(2b) \quad Y_i = \beta_0 + \beta_1 \text{BlackFemale}_i + X_i \beta_2 + e_i$$

Specifically, we first restrict the sample to participants who were randomized into the male treatment condition and use equation 2a. Then, we restrict the sample to participants who were randomized into the female treatment condition and use equation 2b. For both equations, Y is the outcome of interest and i indexes individual teacher respondents. BlackMale_i or BlackFemale_i is an indicator variable equal to 1 if the participant was randomly assigned to a black name. β_1 represents the black-white difference in the outcome of interest. As in equation (1), we include a vector of participant covariates.

² We collected a slightly different set of respondent demographics for Study 1 and 2. The covariates for Study 1 are participant's age, gender, race, teaching experience (indicator variable), and state of residence. The covariates for Study 2 are respondent gender, race, years of teaching, state of residence, and survey duration (in seconds).

Study 1: Troublemaker stereotype

Many adults have negative preconceived notions about youth—that they are self-centered, mischievous, and engage in risky behaviors (Bostrom, 2000; Gilliam & Bales, 2001; Wyn, 2005). Since these prejudices may vary by gender and race, we designed an experimental study to measure participants' views of youths as troublemakers. We predict that more participants will view males as troublemakers than their female counterparts. Additionally, we predict racial gaps within each gender group. Specifically, we predict that more participants will view black males and females as troublemakers compared to their white counterparts.

Method

Participants and Design

Participants were recruited from Amazon's Mechanical Turk (MTurk) program, an online marketplace where employers can hire people to complete small tasks, such as participating in online surveys³. Our initial sample was 512 participants. 52 subjects were excluded due to incomplete surveys or failing the attention checks, leaving a total of 460 individuals. The study is a 2 (race: Black vs. White) X 2 (gender: male vs. female) research design, giving us a total of 4 experimental cells. Participants were randomly assigned into a cell. Table 2.1 presents the summary statistics for Study 1. Column 1 presents descriptive statistics for the full sample, while Columns 2 through 5 displays the average participant characteristics in for each treatment condition (black boy, white girl, black female, or white female). Across the full sample, the average participant age is 36.5, male, and white. Only about 14 percent of our survey sample had any previous teach experience. Additionally, most survey participants were from the Southeast.

³ Many recent studies have found Mturk participants to be as reliable and representative as in-person, college, and data panel sample collection methods. See Buhrmester et al., 2011; Casler et al., 2013, Hauser & Schwarz, 2016; Huff & Tingley, 2015 for more details.

We test for equivalence of the means across all four treatment conditions and present the means in Column 6. There are no statistically significant differences, indicating successful randomization.

Procedure and stimuli

Participants were asked to guess how most Americans view black/white boys/girls across four characteristics: politeness, troublemaking, sassy, and angry. For the purposes of the current study, we focus only on the troublemaking outcome. Participants selected their responses on a scale from 1, *not at all* likely, to 5, *extremely likely*. Next, participants answered a series of background questions—teaching experience, age, gender, race, and state of current residence. Finally, participants were asked to identify the race and gender of the youth group asked about in the survey as an attention check.

Analysis

Tables 2.2 and 2.3 presents our analysis on adult perceptions of youth troublemaking. The “troublemaking” outcome was standardized across the full sample with a mean of 0 and standard deviation of 1. Table 2.2 presents the gender analysis from our regression. The first row of Table 2.2 presents the female average z-score, -0.116 standard deviations. On average, people were less likely to see female youth as troublemakers. After adjusting for covariates, regression estimates suggest a significant 0.215 standard deviation difference in troublemaking perceptions for males. Table 2.3 presents the results for racial differences within each gender group⁴. On average, white males were less likely to be seen as troublemakers (mean z-score of -0.231 standard deviations). Black males were significantly more likely to be perceived as troublemakers (0.732 standard deviation difference). Out of all four treatment groups, white

⁴ Specifically, Column 1 displays the results for equation 2a while Column 2 displays the results for equation 2b.

females were the least likely to be seen as trouble makers (average z-score of -0.494 standard deviations). The racial gap amongst females was a significant 0.711 standard deviations.

Discussion

Consistent with our hypothesis, there are significant differences in adults' preconceived notions of youth as troublemakers. Our results suggest that most Americans view females as less likely to be troublemakers. Within each gender, black youth are more likely to be seen as troublemakers than their white counterparts.

Study 2: Teacher Perceptions

Given the results of Study 1, it is possible that teachers may be carrying prejudices into the classroom, and that these predispositions may color their views of specific student misbehaviors. Study 2 formally tests if teachers' perceptions of student misbehaviors vary by students' gender and, within gender groups, by race. We hypothesize that teachers will be less bothered by females' misbehaviors, and would be less likely to recommend any discipline for these misbehaviors. Similarly, we hypothesize that within each gender group, black students' actions will be seen as more severe and troublesome than those committed by their white counterparts. Finally, we hypothesize that these differential perceptions will increase the likelihood that teachers would recommend discipline for black students compared to their white peers.

Method

Participants and Design

We recruited 284 first-through 12th grade public school teachers from across the United States to participate in our online survey experiment. We restricted our sample to respondents who completed their personal information (geographic location, gender, and race) and answered

all the survey questions. We further drop fifteen individuals who took a long time to answer at least one question⁵. The final sample is 253 individuals, or 89.1% of the original sample.

The study is a 2 (student race: Black vs. White) X 2 (student gender: male vs. female) research design, giving us a total of 4 experimental cells. Participants were randomly assigned into a cell. Table 2.4 presents the summary statistics for our samples. Column 1 presents descriptive statistics for the full sample, while Columns 2 through 5 displays the average teacher characteristics in for each treatment condition (student's name is white male, black male, white female, or black female). Across the full sample, most study participants are white (86%) and female (65%), with the 12.5 years of teaching experience. We test for equivalence of the means across each treatment condition and present the p-value for this hypothesis in Column 6. The only statistically significant difference can be found in teachers' years of experience. Across the four treatment groups, teachers randomized into the black male student treatment had the least amount of teaching experience (mean of 10.6 years) while teachers randomized to receive the black female student name had, on average, the most experience (15.7 years). We control for teaching experience, as well as the other teacher demographics listed in Table 2.4, in all our regressions. Additionally, we test for sensitivity of the results by excluding teaching experience. Our results remain strikingly similar.

Procedure and Stimuli

The four names used in our experiment are Codey (white male), Deonte (black male), Emily (white girl), and LaToya (black girl). We selected these names after extensive testing to ensure that the public perceived that these names were white/black and male/female. One

⁵ We define "long response time" if participants' response time was at or above three standard deviations from the mean for that particular question.

potential worry is that study participants are responding to a perceived socioeconomic status difference rather than a race/gender difference. We address this worry by testing people's perceptions of socioeconomic status attached to each name and ensuring that the names were matched by similar perceived socioeconomic status (see Appendix A for a full description of name testing and selection).

Teachers were presented with a picture of a school and asked to pretend that they are a teacher at this middle school. They are told that all students come from the same middle-income neighborhood, where the typical student-teacher ratio is 27 students to every one teacher. Next, teachers are presented with a disciplinary school record for a student—one scenario involves classroom disruption and one involves insubordination. Scenarios were drawn from actual office referrals from a California school district. We manipulated the order of students' infractions so that half the sample received the classroom disruption scenario first and the other half received the disobedience scenario first (see Appendix B for the text of each scenario). After each scenario, teachers were asked to evaluate how bothered they are by the infraction and how harshly the student should be punished. After completing the questions pertaining to each disciplinary scenario, teachers were asked a series of questions judging their overall views of the student after reading about the students' infractions. Teachers were then told that the experiment portion of the survey is completed and asked to complete some quality assurance questions. Specifically, teachers were probed for suspicion about the study's hypothesis and then asked to select the student's name used in the scenarios from a list of names (to ensure that teachers were paying attention to the student's name). Teachers were given a \$10 Amazon gift card for completing the survey.

Outcome Creation

In total, there are nine outcomes of interest. The first four outcomes explore teachers' reactions to each disciplinary incident. Specifically, teachers were asked "How severe was the student's misbehavior", "To what extent is the student hindering you from maintaining order in your class?", "How irritated do you feel by the student?", and "How severely should the student be disciplined?" Teachers selected their responses on scales ranging from 1, *not at all*, to 5, *extremely*. The severity question aims to get at how serious teachers viewed the infraction. In other words, was the infraction a minor problem that can be easily ignored or did they view the disturbance as a major problem? The second question, about hindrance, aims to measure teachers' perceptions of threat in terms of losing classroom control. Next, we ask about teachers' irritation levels, to see if were particularly bothered by the disciplinary incident. It is possible that even if teachers viewed a student's behavior as a minor disruption, they could be extremely bothered by the action or vice versa. Since each teacher is shown two different disciplinary incidents, teachers answer these first five question twice. All outcomes are standardized by round (first incident, then second incident) across the full sample with a mean of 0 and standard deviation of 1.

The second set of outcomes ask teachers to express their overall opinions of the student, based on these disciplinary infractions. Specifically, we ask teachers to measure the extent to which they think the student is: impolite, a troublemaker, angry, and a danger to other students. Additionally, teachers were asked to judge the extent to which they think the student's behavior is indicative of a pattern. Like the first set of questions, teachers selected their responses on a

scale from 1, *not at all*, to 5, *extremely*. Again, outcomes were standardized across the full sample with a mean of 0 and standard deviation of 1.

Analysis

Gender Analysis

Teachers provided two sets of incident-specific evaluations, one after each disciplinary event. To increase precision, we begin with a model which pooled the estimates⁶ before turning to separated incident-specific analysis. Table 2.5 presents the pooled estimates, with each column representing a separate regression for each outcome—perceived infraction severity (Column 1), hindrance (Column 2), irritation (Column 3), and discipline severity recommendation (Column 4). For ease of comparison, female means are displayed in the first row. There is little evidence of any gender differences when teacher evaluations are pooled across both incidents.

Still, gender differences could exist within each disciplinary event. In fact, Okonofua & Eberhard (2015) found racial differences only after the second disciplinary incident, suggesting that black males are treated more harshly than their white counterparts only after multiple incidents. Figure 1 shows the raw score means, separately by gender and incident. Moving from left to right, the first bar represents the average male score for incident severity after the first incident while the second bar represents the average female incident severity score for the first incident. Correspondingly, the third and fourth bar represent the male and female, respectively, severity score following the second disciplinary infraction. A couple interesting patterns are immediately apparent. First, teachers evaluated the second infraction more harshly than the first,

⁶ Standard errors are clustered at the individual level for the pooled estimation.

suggesting that even small misbehaviors quickly escalate into being perceived as more serious problems. Second, there are some gender differences after the first incident, though these differences decreased or disappeared after the second incident. Teachers appear to view male students' first misbehavior as more of a hindrance and irritation than the same actions when conducted by female students. Additionally, on average, teachers recommended more severe discipline for male students following the first disciplinary infraction than their female counterparts.

It is difficult to tell whether any of the differences in Figure 2.1 are statistically different from one another, so next we turn to our estimating equation (1) and test for significance. Our estimating equation also allows us to increase the precision of our estimates by including controls for teacher demographics and survey duration. Table 2.6 presents these estimates. After controlling for teacher characteristics, there are no statistically significant gender differences for incident severity following either disciplinary incident. Still, teachers reported feeling significantly more hindered (0.283 standard deviations) and irritated (0.258 standard deviations) by males' actions after the first disciplinary incident. While not significant at the traditional 0.05 alpha-level, the point estimate (0.236 standard deviations with a p-value of 0.061) suggests that the increased feelings of hindrance and irritation led some teachers to recommend harsher punishments for male students after the first incident. These gender differences disappeared following the second incident.

We were curious to see if teachers would link these two disciplinary incidents to their overall impression of students. It is possible that teachers could view these disciplinary events as isolated occurrences that represents typical teenage behavior. On the other hand, teachers may

use these incidents to form an overall opinion about the student, and these opinions may vary by students' gender. Consequently, we ask teachers a series of questions to measure how they perceive the students. Table 2.7 presents the results. There is no evidence of gender differences in teachers' perceptions of student characteristics across all five outcomes—impolite, troublemaker, angry, pattern of problematic behavior, and dangerous. While not significant at the traditional 0.05 alpha-value, there are a few estimates with p-values below 0.10 worth noting. Point estimates suggest that teachers were more likely to label male students as troublemakers (0.228 standard deviations with a p-value of 0.082) and angry (0.241 standard deviations with a p-value of 0.055) compared to their female counterparts. Additionally, teachers were more likely to view males as dangerous (0.214 standard deviations with a p-value of 0.089).

Discussion: Gender Analysis

Our results suggest that gender bias exists in teachers' perceptions of student misbehaviors, though in a more limited role than we originally theorized. Overall, teachers were less bothered (in terms of feeling hindered and irritated) by female students' misbehaviors. However, these differences are only significant after the first misbehavior and disappear by the second incident. Furthermore, there were no gender differences in teachers' overall view of students after the second incident. Together, these findings suggest that teachers may be giving female students some leeway when evaluating their misconduct, but that this goodwill is only apparent for the first misbehavior.

Racial Differences Amongst Males

Like the gender analysis, we begin by pooling estimates across both incidents and clustering the standard errors. Regression results are presented in Table 2.8. For ease of

comparison, the white male means are displayed in the first row. There is no evidence of statistical differences in teachers' perceptions of infraction severity. However, teachers felt significantly more hindered (0.279 standard deviations) and irritated (0.511 standard deviations). Although not significant at the traditional 0.05 p-value, point estimates also suggest that teachers were more likely to recommend harsh discipline for black males (0.272 standard deviations, with 0.071 p-value).

Figure 2.2 presents the raw score averages, separated by race and incident occurrence, and provides visual evidence of racial differences amongst males. The first two bars display the average black and white male scores, respectively, for incident severity following the first incident. The third and fourth bars represent the average black and white male scores, respectively, for the second incident's severity. Figure 2.2 suggests a racial bias against black males across all four outcomes—incident severity, hindrance, irritation, and discipline, though it is difficult to tell if these differences are statistically significant. Like Figure 2.1, teachers rated the second incident more harshly than the first incident. In contrast to the gender analysis, which only found differences following the first incident, most of the racial differences amongst male students seem to remain stable. The one exception is teachers' feelings of hindrance. After the first disciplinary incident, teachers felt more hindered by black males' actions than their white counterparts. While this racial difference continues to exist after the second incident, it is dramatically smaller than the racial differences associated with the first misbehavior.

Table 2.9 presents the racial difference amongst males' regression results. Consistent with the patterns found in Figure 2.2, the point estimates suggest racial differences across all four outcomes—though not all are statistically significant. Columns 1 and 2 presents the estimates for

teachers' perceptions on incident severity after the first and second incident, respectively, for male student names. Although there are no statistically significant differences found regarding perceived incident severity, the point estimates (0.166 standard deviations for the first incident and 0.208 standard deviations for the second incident) suggest that teachers view actions by black male students as more severe than when the same were committed by a white male name. After the first disciplinary incident, teachers reported feeling more hindered by black male students (0.404 standard deviations) than their white male counterparts. The racial difference in feeling hindered decreased dramatically after the second incident (0.154 standard deviations) and is no longer statistically significant. Similarly, teachers' irritation of black males decreased from the first incident (0.583 standard deviations) to the second incident (0.440 standard deviations) -- though unlike hindrance, teachers remained significantly more irritated after the second incident. Though not statistically significant, point estimates suggest that teachers were more likely to recommend harsh punishment for black males compared to their white male counterparts (0.270 standard deviations after the first incident and 0.274 after the second incident).

We again examine how teachers view students' overall characteristics after reading about two disciplinary incidents. These results are presented in Table 2.10. On average, teachers more readily assigned negative characteristics to black males. In other words, teachers were more likely to use disciplinary incidents to judge black male students' character. After reviewing the two disciplinary incidents, teachers were significantly more likely to view black males as impolite (0.419 standard deviations) and a troublemaker (0.432 standard deviations). They were also significantly more likely to report that the disciplinary incidents were indicative of a pattern (-0.132 standard deviations). While not significant, the analysis also suggest that teachers viewed

black males as angrier (0.295 standard deviations) and more dangerous (0.256 standard deviations) than their white counterparts. Together, these point estimates suggest that teachers are more likely to excuse away misbehaviors of white male students as isolated incidents while misbehaving black male students are seen as challenging and problematic students.

Discussion: Racial Difference Amongst Males

Overall, these findings suggest a bias against black males in teachers' perceptions of male misbehavior. While the racial differences in hindrance and irritation decrease following the second incident, the irritation estimate remains significant after the second misbehavior, suggesting a slower trajectory towards parity. Furthermore, these two incidents seem to lead to significantly worse overall impressions of black male students. Black males are significantly more likely to be labelled as impolite, a troublemaker, and have their misconducts seen as indicative of a larger pattern of misbehaviors.

Racial Differences Amongst Females

We again start with pooled estimates for teachers' perceptions of female student misbehaviors. Table 2.11 presents these results. There are no statistically significant racial differences across the four incident-specific outcomes: incident severity, hindrance, irritation, and discipline. Aside from discipline (0.225 standard deviations), point estimates also suggest few racial disparities in teachers' perceptions of female actions.

Figure 2.3 presents raw score means of the four incident-specific outcomes, separated by race and incident occurrence (first versus second misbehavior). Contradicting previous group analyses, there appear to be few significant differences for female students even when misbehaviors were separated out by incident. In fact, estimates for black females are remarkably similar to those for white females. Still, there are two potential differences. Although there

appears to be no racial differences in infraction severity, hindrance, and irritation associated with the first infraction, teachers are more likely to recommend harsh punishment for black females. Additionally, and contrary to our hypothesis, teachers seem to view black females' actions as less irritating than white females' actions after the second disciplinary incident. However, it is difficult to tell if any of these differences are statistically significant.

Table 2.12 presents the regression estimates for racial differences amongst females, separated by incident occurrence. Perhaps unsurprising given what we know from Figure 2.3, there are no statistically significant racial differences for female students. While not significant at the traditional 0.05 alpha value (its p-value is 0.070), the point estimate suggests that teachers are more likely to recommend harsh discipline to black females after the first misbehavior (0.367 standard deviations). This difference is surprising given the small (and sometimes negative) point estimates for incident severity, hindrance and irritation and suggests that, at least after the first incident, black females' actions are being differentially penalized compared to their white counterparts. This racial gap quickly disappears by the second disciplinary infraction.

Table 2.13 presents teachers' overall views of female students. Teachers were reluctant to use a few misbehaviors to make judgment calls of white female students' character. On average, teachers did not think poorly of their white female students compared to the full study sample. If anything, the means suggest that teachers were less likely to see white female students as troublemakers (-0.138 standard deviations) or angry (-0.17 standard deviations). Additionally, there is no evidence that teachers had a significantly different view of black females. Point estimates suggest that black females may have been seen as more impolite (0.101 standard deviations), troublemaking (0.170 standard deviations), and angry (0.133 standard deviations)

than their white female counterparts. However, point estimates also suggest teachers may have been less likely to believe that these incidents were part of a pattern of misbehavior (-0.144 standard deviations) and teachers did not see black females as more dangerous than white females (-0.022 standard deviations).

Discussion: Racial Difference Amongst Females

In direct contrast to our hypothesis, we find no evidence of significant racial differences amongst female students. Teachers appear to evaluate female students' misbehaviors in a racially neutral way, although point estimates suggest that they may be more likely to recommend harsh discipline for the first disciplinary incident. Additionally, there is no indication of any differential views of female students' politeness, troublemaking, anger, pattern of problematic behavior, and danger.

Robustness Checks

There are two potential areas of concern regarding study participants' perceptions of the students used in the study. The first possible concern is if the study participants perceived the names as more black or white. If respondents felt that students' names were equally likely to be black or white, we cannot say that any experimental differences we find in the study are due to race. The second area of concern involves participants' perceptions of the student's income status. Ideally, there would be no difference between students' names and perceptions of their income status. If participants believed that specific names connote a low-income status, it would be impossible to tease out whether experimental differences are due to race or income status.

Although both worries were tested out in our pilot study, it is still possible that teachers' perceptions did not match our previous findings. Consequently, our last survey questions asked respondents to evaluate the likelihood that the student presented in the study was 1) black, and 2) from a low-income neighborhood on a scale from 1, *not at all likely*, to 5, *extremely likely*. Appendix Table 2.1 displays respondent's mean responses for these two questions (standardized across the full sample to have a mean of zero and a standard deviation of 1), by treatment condition. Like Tables 1 and 4, the final column presents the p-value for the hypothesis that the means are equal across treatment condition. The first row provides evidence that, on average, study participants viewed students' names as particularly black/white in accordance to their treatment condition. Participants randomized to receive white names were less likely to guess that their student was black, and vice versa. These differences were large and our balance test rejects the null hypothesis that all the means are equivalent at alpha-level 0.05. The second row in Appendix Table 2.1 presents the average guess regarding the likelihood that the student is from a low-income neighborhood. While there is some variation in low-income perceptions (the black female name was the least likely to be from a low-income neighborhood with an average score of -0.1835 while the black male name was the most likely to be from a low-income neighborhood with an average score of 0.1569), we cannot reject the null hypothesis of equivalence at alpha level 0.05. This provides confidence that any experimental differences we find are not due to differential perceptions of income status. As a robustness check, we re-ran our regressions with the inclusion of these two variables—participants' guesses regarding students' race and income. Results are remarkably similar even after controlling for these two variables.

General Discussion and Conclusion

Racial and gender disparities in student discipline has long existed. Researchers and practitioners have long wondered about the role bias may play in how teacher perceive students' actions and subsequent disciplinary decisions. Overall, our study finds that gender/racial biases may play a role in adult perceptions of youth as troublemakers, as well as how teachers assess minor misbehaviors. As a baseline, we find that adults are less likely to view females as troublemakers in comparison to their male counterparts. Additionally, we find that adults are significantly more likely to view black boys and girls as troublemakers than the white counterparts.

In Study 2, we limit our sample to teachers only and measure how teachers view minor student misbehaviors. We did not measure teachers' preconceptions of youth as troublemakers since we did not want to prime participants to think about these preconceptions before presenting them with the disciplinary scenarios. We saw teachers view male actions in a harsher light than their female counterparts after the first disciplinary incident. Interestingly, these differences disappear following the second misbehavior. Together with the results from Study 1, these findings suggest that teachers may be judging students' first misbehaviors based on their preconceptions. Since adults are already more likely to view males as troublemakers, being presented with a misbehavior may serve to confirm their biases, thus warranting a harsher view of male misbehaviors. On the other hand, since girls are less likely to be viewed as troublemakers, their misbehaviors are more likely to be seen as a minor incident and quickly excused. However, this benefit is only extended for the first misbehavior, suggesting that given enough misbehavior incidents, male and female misconduct will be judged equally.

The within gender racial analysis provide some additional insight into the differential role race plays across gender. Beginning with general stereotypes, more adults saw black boys as troublemakers than their white male counterparts. Correspondingly, teachers viewed black male misbehaviors as more problematic (hindering and irritating) than the same misbehaviors by white males. While racial differences decreased by the second incident, irritation remained statistically significant, suggesting a strong bias against black males. Future research should explore whether these racial differences continue decreasing as the number of misbehaviors increase. In other words, is there racial parity after teachers are presented with three or four misbehaviors or do these differences persist?

Despite the perceptions of black males being more hindering and irritating, and contrary to the findings from Okonofua & Eberhardt (2015), we find no evidence that teachers would punish black male students more harshly than their white male counterparts. One possible reason why there was no significant difference in punishment may be because teachers did not view black males' as significantly more serious/dangerous. Another possible factor is the recent rapid change in the culture around school discipline⁷. With the increased federal and local attention on reducing school suspensions, teachers may be more reluctant to assign harsh punishments, particularly for minor infractions. Although teachers did not support harsher punishments for black male students, they were more likely to assign negative characteristics to the same students. Teachers were significantly more likely to label black male students as impolite and a troublemaker. Rather than seeing the disciplinary incidents as a series of unrelated events,

⁷ Since 2010, more than 27 states and 50 of America's largest school districts have created significant changes to their student codes of conduct and discipline policies to include nonpunitive discipline strategies (such as positive behavioral interventions and restorative justice practices) and/or set limits to suspension usage (Eden, 2017; Steinberg and Laco 2017).

teachers were also significantly more likely to think that black male students' actions were indicative of a pattern of problematic behavior. In sum, it appears that racial bias continues to play a role in teachers' perceptions of black male students' actions and judgments of their overall character.

There is little evidence of racial differences amongst female students. There were no significant differences in teacher reports of incident severity, hindrance, and irritation. Teachers appear to see white and black females' actions as equivalent. Yet, our point estimates suggest that, following the first incident, teachers would assign harsher punishments for black females over their white counterparts (0.367 standard deviations, p -value 0.07). Thankfully, this difference disappears by the second incident. Additionally, there are no significant racial differences in teachers' overall impression of students.

Future research should see if the racial discipline gap amongst female can be replicated and if so, why this gap exists. One potential explanation for this gap could be a racial gap in levels of perceived innocence. Youth are generally perceived as still developing and malleable. Consequently, their actions are more likely to be seen as a youthful mistake and they are afforded leniency when disciplinary decisions are made (Giroux, 2000; Steinberg, 2009). However, innocence is not equally afforded to black and white youth. Recent work by Epstein et al. (2017) surveyed a community sample of adults from across the country, and found that black girls were viewed as more independent and less innocent than their white peers. Additionally, these racial gaps began as early as age five.

Chapter 3: Short-term Impacts of Chicago's Suspensions and Expulsions Reduction Plan (SERP)

Introduction

Social scientists, policymakers, and parents have long argued about the merits and costs of out-of-school suspensions (Baker et al, 2001; Browne, 2003; Fabelo et al, 2011; Hemphill et al., 2012; MacGillivray et al, 2008; MacDonald 2012; Monahan et al., 2014; Watanabe and Blume 2015). Suspension proponents argue that disorderly students can create a chaotic learning environment, resulting in learning loss for both themselves and their non-offending classmates. At the same time, critics argue that schools may be over-using suspensions by giving out suspensions for behaviors that could be easily dealt with in school (Advancement Project & The Civil Rights Project, 2000; Baker et al, 2001; Losen & Gillespie, 2012; Losen & Skiba, 2010).

While there is a growing body of literature linking out-of-school suspensions with poor attendance, lower academic achievement, high school dropout, substance abuse, juvenile justice system involvement, and unemployment for suspended students, much of this research is correlational (Browne, 2003; Eaton, 2010; Fabelo et al, 2011; Hemphill et al., 2012; MacGillivray et al, 2008; Monahan et al., 2014; Sweeten, 2006; Wehlage & Rutter, 1986). Despite the lack of causal evidence, the Obama administration made discipline reform policies an education priority, citing studies documenting the disproportionate impact of suspension

policies on students of color, low-income students, and students with disabilities (Losen & Gillespie, 2012; Losen & Skiba, 2010; Skiba et al., 2002; Skiba & Peterson, 1999; Welch & Payne, 2010).

In 2011, the US Departments of Justice and Education created the national Supportive School Discipline Initiative to spearhead discipline-related policy coordination, data collection and research efforts, and provide guidance on alternative discipline (US Department of Justice/Department of Education, 2011). With increased pressure from the federal government and local community members alike, school districts around the country have begun to make changes to reduce suspension use. Today, more than 27 states and 50 of America's largest school districts have created significant changes to their student codes of conduct and discipline policies to include nonpunitive discipline strategies (such as positive behavioral interventions and restorative justice practices) and/or set limits to suspension usage (Eden, 2017; Steinberg and Lacoe 2017). However, there continues to be little causal evidence that suspensions, or suspension reductions, can impact students' outcomes (see Steinberg & Lacoe 2017 for a review). This paper fills the gap in the literature by examining the causal impact of Chicago Public Schools' (CPS) Suspension and Expulsion Reduction Plan (SERP).

Introduced in February 2014, SERP aims to decrease the total number of out-of-school suspensions, increase school accountability for suspension numbers, and expand resources and training about school discipline to staff throughout the district. This ambitious policy change, which caused a precipitous drop in suspension usage, provides a unique opportunity to study the causal effect of out-of-school suspensions on freshmen year outcomes⁸. Following the research

⁸ I focus on freshmen outcomes given freshmen year is a particularly sensitive transition period for students and freshmen year performances are highly predictive of high school graduation (Allensworth & Easton, 2007).

design used by Dee and Jacob (2011), I hypothesize that SERP will have a larger impact on schools that used out-of-school suspensions at higher rates prior to the implementation of SERP. Based on this identifying assumption, I use a difference-in-differences design to estimate the effect of SERP on various outcomes. I identify the effect of SERP by comparing the before-versus-after change in outcomes at historically high suspending schools with the before-versus-after change in outcomes at historically low suspending schools.

First, I estimate the impact of SERP on out-of-suspensions. This analysis shows how quickly, and in what ways, schools in a large public-school system respond to a policy directive from the central office. Having shown that SERP caused a decline in the use of suspensions, I estimate the effect of SERP on other student outcomes, such as attendance, grades, and arrests. Finally, since previous studies have found that students' own academic and behavioral histories can play a role in disciplinary decisions (Bowditch, 1993; Ferguson, 2001; Kinsler, 2011; Morrison et al, 2001; Morrison & D'Incau, 1997; Rocque & Paternoster, 2011), I analyze whether SERP had a differential impact based on students' backgrounds. I calculate individual suspension risk-levels⁹ and explore heterogeneous policy impacts by individual suspension risk-levels.

Overall, SERP significantly reduced out-of-school suspensions, had positive impacts on student attendance and school climate outcomes, and did not adversely impact school-level academic outcomes. Out-of-school suspensions per student declined by 31 percent in the 2013-14 school year (SY14) and by 36.5 percent (compared to the pre-policy mean) during the first full year of implementation (SY15). Additionally, there were no significant changes in the

⁹ Individual suspension risk levels are based on demographic information as well as historical academic and arrest histories.

number of in-school suspensions per student, average in-school suspension lengths, nor the share of students who received at least one in-school suspension for both post-policy years. These results suggest that changes in out-of-school suspension outcomes were not met with an overall shift towards in-school suspensions. Estimates show a significant improvement in attendance in SY14, and no significant impact on academic achievement. Contrary to public fears that reducing suspensions would create more chaos in the classroom (MacDonald 2012; Watanabe and Blume 2015), students reported significantly *more* classroom order in SY15, and no changes in other measures of school climate. Additionally, there were no detectable changes to school-level arrests per student.

While school-level analysis provides a good overview of SERP's general impact, a look at heterogeneous impacts sheds more light on who is benefitting from SERP and in what ways. Restricting the analysis by individual-level risk group allows me to separately estimate the policy impact on high risk students and then lower risk students.

SERP had a bit of a complicated impact on high risk students. There was little impact on high risk students' out of school suspensions for SY14, though point estimates suggest a slight decrease. By the first full year of implementation, SY15, high risk students experienced a 49.6% decline in out of school suspension incidents. Analysis also shows that in addition to fewer out-of-school suspensions, SERP also impacted *who* was being suspended. Compared to the pre-policy mean, 24.7% *fewer* students received a suspension during SY15. Additionally, there were no significant changes in in-school suspensions, though point estimates suggest that in-school suspensions may have increased after SERP introduction.

Mirroring findings from the school-level analysis, there is little evidence from the student-level analysis that out-of-school suspensions were met with much harm. There were no significant changes in attendance, achievement, or school climate reports. To be sure, student reports on classroom order suggest a bumpy policy transition. Significantly fewer students reported orderly classrooms in SY14. However, by SY15, student reports on classroom order were indistinguishable to the pre-policy mean. Furthermore, arrest charges among high-risk students declined by almost 23% (from the pre-policy mean) in SY15.

Lower risk students also benefitted from SERP. On average, lower risk students received 0.412 fewer out-of-school suspensions during SY15, representing a 64% decline from the pre-policy mean. Correspondingly, the average out-of-school suspension length was reduced by about one day (a 60.5% decline from the pre-policy mean) and the share of students who received at least one out-of-school suspension fell by 16.8 percentage points (or 53.5% drop compared to the pre-policy mean). Attendance also improved (significantly in SY15), and there is no evidence of harm to freshmen achievement, classroom order, or school climate outcomes. Importantly, SERP led to significantly fewer arrest charges (-.060 fewer arrests, or a 63% reduction).

This study extends the existing literature in at least five ways. First, most of the literature on suspension reduction policies are limited to descriptive analysis (Eden 2017; Stevens et al, 2015). These studies found an association between suspension reduction policies and deteriorated school climate, but do not establish a causal relationship. Lacoe and Steinberg (in press) brings the field closer to a causal understanding of suspension policies by using a difference-in-differences approach and comparing district-level changes in outcomes for

Philadelphia with changes in outcomes for districts in the rest of the state. However, the policy ultimately did not bring about any significant changes to out-of-school suspensions. My study provides the opportunity to estimate the causal impact of a disciplinary policy reform that significantly reduced out-of-school suspensions. Second, by constructing a comparison group within the same urban school district, I decrease the possibility that changes in outcomes are due to other environmental and policy factors that could be occurring at the same time as SERP introduction. Third, a recent study by Aizer and Doyle, Jr. (2015) show that juvenile incarceration significantly decreases high school completion and increases adult incarceration rates. While I do not have incarceration records, I am able to link students to arrest charges. The availability of Chicago Policy Department data allows me to analyze the causal impact of a suspension reduction policy on arrests. Fourth, individual-level longitudinal data allow me to test the parallel trends assumption and to control for shifts in student composition across the study period. Finally, this study leverages the individual level data to provide insight on how individual suspension risk levels may moderate disciplinary policies' impact on student outcomes.

The remainder of the paper is as follows. Section 2 details Chicago's suspension reduction policy, Section 3 presents the theoretical framework. Section 4 describes the data and Section 5 presents the research design. Section 6 presents the results for the school-level analysis and Section 7 takes a closer look at heterogeneous effects by analyzing SERP's impact on individual-level risk groups. Section 8 provides a robustness check and Section 9 concludes.

Suspension and Expulsion Reduction Plan

Since 2008, The Chicago Public Schools (CPS) has recognized the growing literature linking exclusionary discipline policies with poor student outcomes and is actively working to

decrease schools' suspension usage. Introduced in February 2014, the Suspension and Expulsion Reduction Plan (SERP) is CPS' attempt at a large-scale disciplinary policy reform. These policies limited suspension usage in several ways and was formally codified in the June 2014 revision of the Student Code of Conduct¹⁰. I highlight a few of the biggest changes here (see Appendix C for a more detailed description of the Student Code of Conduct changes). First, schools must receive central office approval for suspensions in pre-kindergarten through second grade. Second, out-of-school suspensions usage is more limited. Infractions are grouped into six different categories with Group 1 infractions representing inappropriate behaviors and Group 6 representing illegal or seriously disruptive behaviors. Each group has an assigned list of possible consequences and interventions. Behavioral infractions in Groups 1 through 3 are considered minor infractions while infractions in Groups 4 through 6 are considered major infractions. SERP eliminated out-of-school suspension as a possible consequence for all Group 2 behaviors and for first-time Group 3 infractions. Furthermore, out-of-school suspensions are permitted for students in grades three through twelve "only if students' attendance endangers others, student causes chronic or extreme interruption to other students' participation in school activities, and prior interventions have been utilized" (Chicago Board of Education, 2014). Administrators must develop a support plan for students who are suspended for three or more days. SERP also aims to limit the number of suspension days: the number of suspension days permitted for Group 3 through 5 infractions is lower and schools now have discretion to assign shorter suspension terms for Group 5 and 6 infractions.

¹⁰ Formal changes to the Student Code of Conduct occur once a year, usually in June or July. SERP began in February. Starting in February, schools and principals were made aware of the policy, that there will be increased accountability for school-level suspension numbers, and what types of support CPS was providing for principals, teachers, and students.

These directives were paired with extra training and supports for schools in their transition to alternative discipline strategies. Central office advocates the use of the Multi-Tiered System of Supports (MTSS), which seeks to provide instructional and resource support for schools at three levels: all students, students with moderate behavior problems, and students in need of intensive support (CPS 2016). Over 200 support staff have been added to the District to support MTSS interventions such as restorative practice coaches, climate support teams, and classroom management coaches. Additionally, CPS has created a team of Social-Emotional Learning (SEL) specialists who provide targeted support tailored to schools' unique needs within each network group (Chicago Public Schools, 2015).

Theoretical Framework

Suspension policies can be thought to impact three groups of students: students who will always be suspended, students who may or may not be suspended depending on policy changes (I call this group the marginally misbehaving students), and then students who will never be suspended. An evaluation of a suspension reduction policy analyzes how the policy impacts these latter two student groups. First, it measures the effect of not being suspended for marginally misbehaving students. Second, it measures the peer effect of keeping marginally misbehaving students in school.

Developmental psychologists underscore the importance of environments in shaping youth development (Bronfenbrenner 1986; Bronfenbrenner & Morris 1998; Sameroff 1983, 1994; Sameroff & Fiese, 2000). In other words, *how* schools interact with students' developmentally normative misbehaviors can play an important role in their outcomes. Psychologists theorizes that schools that rely heavily on suspensions may put marginally

misbehaving students on the path of academic failure. Students who receive an overly harsh punishment, especially for a first or minor offense, may interpret the disciplinary decision as a rejection, and begin to question the importance of school to their identity (Roderick, 2003). If teens perceive schools as a place of constant rejection, they may disengage from school.

An overly punishing environment could also be harmful for students who would never be suspended. These students can begin to distrust the school's judgment and legitimacy, creating an anxious environment and making it harder for students to succeed in school (Kirk and Papachristos 2011; Kupchik 2010; Morris 2005; Nolan 2011). In a longitudinal study of *non-suspended* students from a large Kentucky metropolitan school district, Perry & Morris (2014) found higher test scores for in semesters where schools assigned fewer suspensions compared to semesters with higher suspension levels. These results remained statistically significant even after controlling for schools' overall levels of discipline and disorganization. Consequently, one hypothesis is that as schools move towards alternative discipline strategies, particularly ones that allow youth to learn from their mistakes and develop their social-emotional skills, students (both behaviorally marginal students and their non-offending classmates) may benefit. These positive interventions could create stronger bonds to school and improve outcomes for all students.

An alternative theory suggests that limiting suspension usage can be detrimental for all students. As disruptive students realize that out-of-school suspensions will rarely be used, they may be encouraged to act out even more in class. Teachers may be forced to spend more of their time dealing with classroom distractions, leading to a loss in learning time for all students. Empirical studies have found that exposure to disruptive students increased non-offending students' misbehaviors while lowering test scores and attendance (Carrell & Hoekstra 2010;

Imberman et al 2012). As classrooms become more chaotic, non-offending kids may start avoiding class either due to reasons such as safety concerns or a simple wish to avoid classroom disorder (Bowen & Bowen 1999; Bryk 2010). A recent study by Carrell, Hoekstra, & Kuka (2016) find that having a disruptive classmate in elementary school also has a significant long-term impact: age 26 earnings were reduced by 3 to 4 percent.

Data

Data for this project comes from a variety of different sources. Longitudinal individual-level data comes from the Chicago Public School. The main outcome of interest is “number of out-of-school suspensions per student” (total number of suspensions divided by total student enrollment). This measure of suspensions helps to account for changing student populations at each school and across schools. Data for suspension outcomes are pulled from CPS misconduct files, available from School Year 2010-2011 (SY11) through School Year 2014-2015 (SY15). Supplemental suspension information is drawn from CPS attendance files available from SY09 through SY15 to measure average length of suspensions and an indicator for if a student was ever suspended.

Given the richness of the available data, it is possible to test numerous outcomes. However, doing so would also increase the risk of false positives (Anderson 2008; Kling, Liebman, and Katz 2007). As an alternative, I create three summary indices for closely related outcomes: attendance, academic, and school climate. Indices help to increase the statistical power to detect effects of variables that highly correlated with one another. Each of these indices are z-scored by the average pre-policy mean and standard deviations. Variables with a negative

valence are reverse coded before they are z-scored. Consequently, higher index scores represent improvements in attendance, academic outcomes, or school climate.

Attendance and academic outcomes are often seen as a measure of a student's engagement level; a combination of lower grades and higher absence rates are seen to represent students who are mostly disconnected from schools. The composite attendance index includes present days, unexcused absences (reverse-coded), and excused absences (reverse-coded). The composite academic index includes grade point average for all core classes, year-end accumulated credit, and the freshman on-track indicator¹¹. Researchers at the University of Chicago Consortium on School Research have developed the "freshman on-track indicator," from Chicago Public Schools data, to predict who is likely to graduate or dropout based on freshman year course performance (Allensworth & Easton, 2007). Students are defined as "on-track" if they accumulate at least ten semester credits (five full year credits) and have no more than one F in a core course (defined as English, math, science, or social science). "On-track" students are four times more likely to graduate high school than their off-track counterparts (Allensworth & Easton, 2005).

To complement CPS's administrative data, I pull in responses from the annual CPS student survey My Voice My School to create the *classroom order* indicator and school climate index¹². I separate out *classroom order* from the rest of the school climate index since previous literature suggests a link between suspension reduction policies and classroom chaos. *Classroom order* measures the extent to which students view their classroom as out-of-control (this measure

¹¹ Due to some unforeseen circumstances, CPS test scores are not available for SY15. Consequently, I am not able to include test scores in my academic index.

¹² Ideally, I would analyze both student- and teacher-reports on classroom order and school climate. Unfortunately, I only have access to student survey responses.

is reverse coded). However, school climates can change in ways beyond classroom disorder. For example, it is possible that there will be more bullying or more concerns about school safety after SERP. Alternatively, an increase in alternative consequences that involved building stronger relationships could increase students' feelings of being connected to teachers and the school. The composite school climate index is created by averaging the z-score for student reports on bullying (reverse coded), positive peer environment, personal safety, school safety concerns (reverse coded), student-teacher connection, and school-connectedness. (See Appendix D for exact survey question wording)

Finally, the Chicago Police Department has provided me with data on student arrest. I hypothesize that SERP may have had impact on arrest charges and in the interest of having fewer outcomes, I only use the outcome "total arrest charges." It is possible that if schools in the post-SERP era feel like their hands are tied in terms on reprimanding students, they may be *more* likely to call police to handle simple misbehaviors (Mendis 2017; Watanabe 2015). On the other hand, if SERP increases a sense of school community, it is possible that schools are more likely to handle misdemeanor offenses (such as fighting) in-house rather than contacting police.

Empirical Strategy

The causal impact of suspension policies on student outcomes has been difficult to determine. One way to obtain a causal estimate of suspension policies on student outcomes would be to randomly assign schools to allow suspensions at some schools and ban suspensions at others. Student outcomes such as attendance, grade point averages, arrests, etc. would be tracked over time. Any differences in short- and long-term outcomes could then be causally assigned to suspension policies. However, most school districts would not agree to random

assignment of disciplinary policies for feasibility and ethical reasons. Even if there was a school district that would be willing to experiment with disciplinary policies, it would be difficult to imagine that school disciplinarians would follow through with the randomization results and either suspend students or not based on an experimental study.

Since randomization is not practical, one solution is to use SERP to estimate the causal impact of a suspension reduction policy on students' outcomes. I use a difference-in-differences approach to leverage variation over time and across schools. I begin by comparing pre-SERP and post-SERP differences in outcomes. However, a simple pre-post comparison does not quite get us to a causal understanding of suspension reduction policies since there could have been many other policy changes at the time, or there could have been a natural change in outcomes that would have occurred even without SERP. SERP was implemented district-wide all at once, so an additional challenge is finding a comparison group of schools. To generate this comparison group, I follow Dee and Jacob (2011) and rely on the assumption that policies will have a stronger impact on some sub-populations and little or no impact on others. Specifically, I hypothesize that SERP, the "treatment", should have a larger impact on schools that used out of school suspensions at higher rates prior to the implementation of SERP. Similarly, it should have a larger impact on students that were more likely to be suspended under pre-SERP policies (i.e. students with a history of suspensions, low test scores, etc.) than students who were less likely to be suspended before SERP. In contrast, I hypothesize that historically low-suspending schools and students with a lower risk of being suspended will not be as affected by SERP. I estimate two types of difference-in-differences models: a school-level analysis that uses historically high suspending schools as a treatment group and historically low suspending schools as the

comparison group, and a student-level analysis that restricts the sample within treatment and comparison schools to high-risk students and then lower-risk students.

At its most basic form, the difference-in-difference estimator presents three different subtractions: (A) pre-post change in an outcome for historically high suspending schools, (B) pre-post change in an outcome for historically low suspending schools, and (A-B) the difference between these two changes. In brief, the difference-in-differences estimate is the before-after policy difference in outcomes in the high BSR group *relative* to the before-after policy difference in outcomes for the low BSR group. I first use the difference-in-differences model to measure the effect of the suspension-reduction policy (SERP) on the rate at which schools suspend students, both out-of-school suspensions and in-school suspensions. This estimate shows how a policy directive from CPS central administration effected disciplinary practice at schools. I then use the difference-in-differences model to measure the effect of reducing out-of-school suspensions on student outcomes, including attendance, academic achievement, school climate, and arrests.

Focus on Freshmen Year Outcomes

Freshmen year is a sensitive transition for many youth and brings stress to many young students who worry about bullying, a greater amount of homework, class difficulty, and building relationships during the move to a new environment (Akos & Galassi, 2004; Allensworth & Easton, 2005; Zeedyk et al., 2003). It is unsurprising that ninth graders report feeling lower levels of school attachment/engagement and higher levels of depression symptoms (Barber & Olsen, 2004; Newman et al, 2007; Seidman et al, 1996) than students in earlier grades. Freshman year is also often accompanied by declines in grade point averages (GPAs) and attendance records (Benner, 2011; Felner, Primavera, & Cauce, 1981; Schwerdt & West, 2013; Seidman et al, 1996). These findings are distressing given that freshman year course is more predictive of

high school graduation than elementary school test scores, grade retention, and background demographics such as gender, race, and socioeconomic status (Allensworth & Easton, 2007).

While students are trying to navigate their way through freshmen year, schools are trying to socialize these new students into following certain behavioral expectations. One approach schools take to regulate behavior is through out-of-school suspensions. Figure 3.1 presents a graphic description of how out-of-school suspension rates (defined as total number of suspensions divided by total number of students) vary across grade levels and school years 2010-2011 through 2014-2015. As expected, most out-of-school suspensions occur in 9th and 10th grade, with a decline in suspension rates for 11th grade. 12th grade suspension rates, at about 50% of the freshmen/sophomore suspension rates, are the lowest of all four grade levels. Consequently, SERP, a policy intended to significantly decrease out-of-school suspensions, should have the largest impact on freshmen and sophomores.

Given the dramatic changes that can occur during freshmen year, and the large amount of out-of-school suspensions that occur during freshmen year, I restrict all data analysis to freshmen year outcomes. Limiting analysis to freshmen year outcomes also brings two additional benefits. First, freshmen are less likely to drop out of school. Illinois law requires students to be enrolled until their 17th birthday. Since most freshmen are not yet 17, there is less worry that students will drop out and bias the sample. For example, if only high achieving students are staying in school through senior year, it would be difficult to determine if better outcomes in senior year are due to a policy change or simply due to a change in the analysis sample. Second, academic performance during freshmen year is highly predictive of high school completion¹³.

¹³ Ideally, I would test SERP impact on longer term outcomes such as high school drop-out and high school graduation. Unfortunately, I only have access to two years of post-policy data.

School-Level Analysis

The school-level difference-in-differences analysis leverages the variation in suspension usage that existed in schools across CPS in the pre-policy period. I create the historically high suspending school group through a multi-step process. First, I use individual-level CPS data and assign students to the school they were enrolled in at the beginning of the school year (CPS uses the 20th day of school). The data is then collapsed at the school-year level to create counts of annual out-of-school suspension per student (total number of out-of-school suspensions/total number of students). Since high numbers of out-of-school suspensions can be due to either many suspensions or a large student population, “suspensions per student” is the preferred outcome of interest. Annual suspensions per student are then averaged across the three pre-policy school years (school years 2010-2011, 2011-2012, and 2012-2013). The sample is restricted to mainstream public high schools. Schools with a three-year average of out-of-school suspensions that are at or above the 75th percentile (0.73 suspensions per student) are defined as high suspending, or treatment, schools while schools with a three-year average at or below the 25th percentile (0.24 suspensions per student) are low suspending, or comparison, schools.

Model

Data is first restricted to freshmen students only and then collapsed to the school-year level for all school-level analysis. Taking school-level averages helps to account for varying population sizes across the schools and provides a conservative way to account for school-year level error clustering. Each school has five years of data. I estimate the difference-in-difference model with school and year fixed effects which is described by the following equation (1):

$$(1) Y_{st} = \beta_0 + \pi_t + \gamma_s + \beta_1(\text{high suspending} * \text{SY14})_{st} + \beta_2(\text{high suspending} * \text{SY15})_{st} + X_{st} \beta_3 + e_{st}$$

where Y is the outcome of interest, t indexes year, and s indexes school. The matrix π includes year fixed effects and the matrix γ contains school fixed effects. Since SERP was introduced in the middle of SY14, it is possible that the policy had a different impact on students' outcomes for SY14 than in SY15. I allow for this possibility by creating separate estimates for each post-policy year (high suspending school*SY14, high suspending school*SY15). X represents a matrix of time-varying school-level covariates that may influence a school's suspension level (such as share of the freshmen cohort that is male or has an Individualized Education Program, or IEP). e_{st} represents the error term. To address serial correlation concerns, standard errors are clustered at the school level.

β_0 estimates the outcome variable for low suspending schools in the pre-policy period. β_1 estimates the high versus low suspending schools difference in SY14, while β_2 captures the difference between high and low suspending schools in SY15.

Identification in this model relies on the assumption that, in the absence of SERP, high and low suspending schools would have similar trends in outcomes. I formally test this assumption by amending equation (1) to include interactions between year fixed effects and the “high suspending school” (treatment) indicator. Specifically, I estimate the parallel trends assumption with equation (2):

$$(2) Y_{st} = \beta_0 + \pi_t + \gamma_s + \beta_1(\text{high suspending} * \text{SY11})_{st} + \beta_2(\text{high suspending} * \text{SY12})_{st} + \beta_3(\text{high suspending} * \text{SY14})_{st} + \beta_4(\text{high suspending} * \text{SY15})_{st} + X_{st} \beta_5 + e_{st}$$

where Y is the outcome of interest for school s at time t . Matrix π includes year fixed effects and the matrix γ contains school fixed effects. X represents a matrix of time-varying school-level covariates and e_{st} represents the error term. Standard errors are clustered at the school level.

The “high suspending school*year” interaction measures the high vs low-suspending school difference for each year. Additionally, I leave out the high suspending school*SY13 interaction and use it as the reference year¹⁴. Consequently, β_1 through β_4 are interpreted as the additional high- versus low-suspending difference for that school year in comparison to the high- versus low-suspending difference in SY13. In other words, β_1 estimates the change between high and low suspending school differences for SY11 in comparison to the difference found in SY13. Correspondingly, β_2 , β_3 , and β_4 estimates the changes during SY12, SY14, and SY15, respectively, compared to the SY13 difference. If the parallel trends assumption holds, there should be no significant estimates for β_1 and β_2 .

Student-Level Analysis

Suspension experiences vary drastically even *within* a school (Ferguson, 2001; Morrison & D'Incau, 1997; Vavrus & Cole, 2002). Some students are more likely to be suspended than others based on their individual characteristics and school history. This analysis uses students' individual suspension likelihood to evaluate if SERP has heterogeneous impacts based on combinations of individual- and school-level suspension risk. It is possible that SERP was particularly beneficial to certain groups of students rather than all students. It can be helpful to both policymakers and practitioners to understand how SERP may differentially impact students within both high- and low- suspending schools.

¹⁴ I leave out SY13 and use it as the reference year since it is the last full school year before SERP.

To generate individual suspension likelihood, I use a series of demographic (race, gender, IEP, free/reduced lunch status), academic (attendance rates, grades, and test scores in 7th and 8th grade), and arrest history variables to create a prediction score of who is most likely to receive at least one out-of-school suspension during freshmen year. See Appendix C for more details on the estimation and for the regression coefficients from the regression. The prediction score is then used to create a group of high suspension risk students. I assign students as high risk if their prediction score is at or above the 75th percentile (0.251). The rest of the students are assigned as lower risk students.

With some slight modifications to equation (1), the student-level analysis is described as follows:

$$(3) Y_{ist} = \beta_0 + \pi_t + \gamma_s + \beta_1(\text{high suspending} * \text{SY14})_{st} + \beta_2(\text{high suspending} * \text{SY15})_{st} + X_{ist} \beta_4 + e_{ist}$$

where Y is the outcome of interest, t indexes year, and s indexes student. The matrix π includes year fixed effects and the matrix γ contains school fixed effects. Additionally, “X” represents a matrix of individual-level covariates that may influence a student’s likelihood of being suspended (the same covariates used to create the prediction score such as male, has an IEP, etc. and can be found in Appendix C). I begin by using equation (3) and restricting the sample to, high risk students and then repeating the process and restricting the sample to lower risk students. Corresponding to the school-level analysis, students attending high suspending schools are the treatment group while students attending low suspending schools are the comparison group.

As with the school-level analysis, identification in this model relies on the assumption that, in the absence of SERP, students at high suspending and low suspending schools would have similar trends in outcomes. I formally test this assumption with equation (4):

$$(4) \ Y_{ist} = \beta_0 + \pi_t + \gamma_s + \beta_1(\text{high suspending} * \text{SY11})_{st} + \beta_2(\text{high suspending} * \text{SY12})_{st} + \beta_3(\text{high suspending} * \text{SY14})_{st} + \beta_4(\text{high suspending} * \text{SY15})_{st} + X_{ist} \beta_5 + e_{ist}$$

where Y is the outcome of interest for student s at time t . Matrix π includes year fixed effects and the matrix γ contains school fixed effects. “ X ” represents a matrix of individual-level covariates found in Appendix C and e_{st} represents the error term. Standard errors are clustered at the school level.

School Level Results

Research Question 1: What is SERP’s impact on school-level out-of-school suspensions?

As previously discussed, Figure 3.1 measures the trend in out-of-school suspensions per student over grade levels and school years. It also provides preliminary evidence that out-of-school suspensions per student decreased in the post-policy years—particularly during SY15. Regression analysis allows me to estimate SERP impact on out-of-school suspension (OSS) per student separately for SY14 and SY15 while also adding in a series of different controls. Table 3.1 presents estimates for OSS per student (SERP’s main goal), and how estimates vary across different econometric specifications. The β s of interest are “high suspending school*SY14” and “high suspending school *SY15.” These coefficients estimate the additional post-policy change in outcomes, OSS per student in this case, for high suspending schools (compared to low suspending schools) during SY14 and SY15 respectively. Column 1 provides the estimates for the simplest difference-in-differences estimate, where no controls and fixed effects are included. According to this model, high suspending schools had a 0.185 decline in OSS per student rate during SY14. This change is not significantly different from the decline rate at low suspending

schools. In other words, there were no significant impacts on SY14 suspension rates. However, the story changes in SY15; SERP caused a 0.697 decline (or 51.3% drop from their pre-policy mean) in out-of-school suspensions per student during its first full year of implementation.

Column 2 of Table 3.1 adds in time-varying school-level covariates and year fixed effects. The time-varying school-level covariates help adjust estimates for any shifts in student characteristics that may be correlated with the outcome. These covariates include: the share of student population that are male, speaks a language other than English at home, black, Hispanic, white, had an IEP, designated as Emotionally Disturbed, and received free/reduced lunch. It also includes the total student population size as well as a series of 8th grade characteristics for the freshmen cohort: reading test score, math test score, attendance rate, and share of students with any arrest charges during 8th grade¹⁵. Year-fixed effects are added to control for any individual year shocks that impacted all schools and could be correlated with the outcome. With the addition of covariates and year fixed effects, there continues to be no statistically significant impacts during SY14. Additionally, the SY15 SERP estimate decreases from -0.697 to -0.593, but remains statistically significant.

The specification in column 3 includes everything used in column 2 (school-level covariates and year fixed effects) and adds in school-fixed effects. This is the preferred estimate since school-fixed effects also controls for all time-invariant variables unique to each school that may be correlated with the outcome. With the addition of school-fixed effects, there is now a significant decrease in out-of-school suspension rates for SY14. During this first year of

¹⁵ One potential concern is that SERP could change students' 8th grade characteristics for the second cohort (they experience SERP halfway through their 8th grade year). I run robustness checks by excluding these 8th grade characteristics from the covariates and find remarkably similar results.

implementation, freshmen experienced a 0.421 drop in their suspension rates, or a 31.0% change from their pre-policy mean. Additionally, the estimated SERP effect on SY15 out-of-school suspension rates is a more conservative (compared to previous models) -0.495, representing a 36.5% decline relative to the pre-policy mean for high BSR schools. For the sake of brevity, I only present results from my preferred specification (shown in column 3 and includes school-level covariates as well as year and school fixed estimates) for the rest of the outcomes¹⁶.

One possible concern is that the regression results are due to differences in pre-policy trends. For example, if out-of-school suspensions were already declining at a faster rate in high suspending schools, it would be difficult to tell if out-of-school suspension rates were declining due to SERP or due to pre-existing trends. I use equation (2) to formally test this concern, and present the results in Appendix Table 3.2. For visual ease, I plot the year by high suspending school dummy interactions in Figure 3.2A¹⁷. The figure suggests that there were no dramatic pre-trends in before SERP (SY11 through SY13). If anything, the figure suggests that out of school suspension rates may have been increasing at a slightly faster rate in high suspending schools compared to low suspending schools during the pre-policy period. Furthermore, there is large dip in out-of-school suspension rates starting in SY14, suggesting that SERP was successful in decreasing out-of-school suspension rates. I repeat the same process for every outcome, but for brevity, only show the graphs for “share of students who have received at least one out-of-school suspension during the school year” (Figure 3.2B), in-school suspensions per student (Figure 3.2C), and “share of students who have received at least one in-school suspension during the

¹⁶ For consistency across all three models, I present the r-squared for the “between” model in Table 3.1, Column 3. However, I follow custom and report the r-squared for the “within” model for the remaining regressions.

¹⁷In order to plot all five school years, I run equation (2), but leave out the school fixed effects.

school year” (Figure 3.2D). Again, Figure 3.2B suggests that there are no worrisome pre-trends and provides evidence that SERP decreased the proportion of students who received an out-of-school suspension. Figures 3.2C and 3.2D show that there is high variation in schools’ in-school suspension usage, and that high-suspending schools may have been increasing their use of in-school suspensions during the pre-policy years. It is difficult to tell, visually, if this increasing pattern is statistically significant or if there are significant changes to the trend in the post-policy years. However, the regressions results in Appendix Table 3.2 suggest no worrisome pre-trends for in-school suspension outcomes.

Table 3.2 presents the regression estimates for all the suspension outcomes. In addition to out-of-school suspensions per student, I hypothesize that SERP may have also impacted the average number of OSS days and the share of students who have received at least one OSS during the school year. There are two measures of OSS lengths—one from the misconduct file and one from the attendance file. Schools record suspension information in two different databases. Ideally, these two data sources would result in an identical match, but this is not the case. Some suspension lengths are only recorded in the misconduct file while others are only recorded in the attendance file. I only present results from the misconduct data file, but results using the attendance file are remarkably similar and included in Appendix Table 3.3.

Columns 1 through 3 suggest that SERP brought some immediate improvements to out-of-school suspension outcomes, and even larger improvement during the first full year of implementation (SY15). As discussed in Table 3.1, there was a significant decline in out-of-school suspension rates—a 31.0% decline in SY14 and a 36.5% decline in SY15 relative to the pre-policy average. This decline was also met with fewer average out of school suspension days.

Misconduct file data estimates a decline in average OSS days by 1.629 days (37.4% decrease from the pre-policy mean) in SY14 and 2.522 days (57.9% decrease from the pre-policy mean) in SY15. It is possible that SERP simply cut down the number of suspensions, with no impact on the share of students who received an out-of-school suspension. The estimates in column 3 suggest that this was not the case. SERP significantly changed the population of students who were receiving OSS. The share of suspended students dropped by 9.1 percentage points (representing a 18.2% decline from the pre-policy mean) in SY14 and 12.6 percentage points (or 25.2% decrease from the pre-policy average) for SY15.

With the large decline in out of school suspensions, some worry that schools may simply replace their OSS consequences with in-school suspensions (ISS). The estimates in Columns 4-6 suggest that this is not the case. SERP did not significantly increase the in-school suspension rate, the average number of in-school suspension days, nor the share of students with at least one ISS for both post-policy years. The standard errors for in-school suspension outcomes are quite large for SY15. Consequently, it would be impossible to rule out either a 0.458 decrease or a 2.414 increase in ISS suspensions per students for SY15. On average, both high- and low-suspending schools have been slowly increasing their usage of in-school suspensions across the study period (SY11-SY15). However, standard errors also increased. In other words, there was also more variation in in-school suspension usage across the study period. Qualitative interviews suggest that one potential reason may be due to funding fluctuations; some schools did not have the funding to staff a consistent in-school suspension room while other schools were consistently able to use in-school suspensions as a consequence. Taken together, these results indicate that

SERP was successful in decreasing schools' out-of-school suspension usage, without significantly increasing in-school suspension usage.

Given the national attention on racial and gender discipline disparities, it is also important to analyze how SERP may have differentially impacted students across racial and gender groups. Table 3.3 presents the estimates for share of students who received at least one out-of-school suspension during the school year by race and gender. Estimates for pre-policy averages at historically high suspending schools (displayed in the first row of data) are consistent with what has been found in the existing suspension literature—black males were most likely to receive an OSS, with black females following close behind. White males and females are the least likely to be suspended, with Hispanic males and females found in between the two racial groups. Columns 1 through 6 present SERP's impact on out-of-school suspensions across various racial and gender subpopulations. There are no statistically significant changes in the share of subpopulations who received at least one out of school suspension.

To more directly measure suspension changes relative to different racial groups, I create a measure of the suspension risk gap. The suspension risk gap is calculated by taking the share of students with at least one OSS from one racial group and subtracting the share of students with at least one OSS from their white counterparts. For example, Column 7 takes the share of black males who received at least one OSS and subtracts the share of white males who received at least one OSS during the same year. The regression coefficients then estimate whether this gap is increasing or decreasing at a significantly different rate than compared to low suspending schools. If suspensions rates were equally distributed across each racial group, the suspension risk gap will equal to zero. Regression results suggest a decreasing black/white male gap (33%

decline in SY14 and 38% decline in SY15 compared to the pre-policy mean). Estimates in Columns 1 and 5 suggest that this significant finding is due to both a decline the black male suspension rate and a slight increase in the white male suspension rate. The estimates also suggest a movement towards racial parity for black females, Hispanic males, and Hispanic females during SY15, though note that none of these estimates are statistically significant.

Research Question 2: What is SERP's impact on other school and arrest outcomes?

Suspension reduction policy opponents argue that such policies will lead to chaos in the classroom and a loss in school engagement for all students. Table 3.4 presents evidence that these fears did not materialize for CPS students in this study. Again, the first row of data presents the pre-policy mean for high-suspending schools across each z-scored index. High-suspending schools had lower than average attendance, academic, classroom order, and school climate scores during the pre-policy years. SERP significantly increased the attendance index score by 0.162 standard deviations (a 60.7% increase from the pre-policy mean) in SY14. While the SY15 estimate remain positive (0.105 standard deviation increase, or a 39% improvement from the pre-policy mean), it is no longer statistically significant. Opponents of suspension reduction policies worry that keeping disruptive students in school can damage overall school achievement. I find no evidence of harm to students' progression towards high school completion. Regression estimates presented in Column 2 suggest that the significant decrease in out-of-school suspensions were not met with any significant changes in freshmen year academic achievement.

Columns 3 and 4 present the regression estimates for school climate survey responses. Despite public outcry that restricting suspensions would increase disorder in the classroom (Klein, 2014; Perez, 2015), student reports reflect a surprising amount of stability, and even a bit of improvement, in school climate. Students were more likely to describe their classrooms as

well-ordered in both post-policy years, though the change was only significant during SY15 (an increase of 0.407 standard deviations, or an 106% improvement from the pre-policy mean). Student reports on school climate also suggest quite a bit of stability outside the classroom. There are no significant changes in the school climate index which, as a reminder, includes factors such as bullying, peers, personal and school safety, as well as connections to teachers and school.

One area of potential concern regarding the school survey responses is its high non-response rates—about 44.5% of high school students in the sample did not complete the school survey. However, analysis of the missing patterns show that survey response was remarkably stable across various demographic characteristics during the study period and across both high and low suspending schools. Appendix Table 3.4 provides detailed estimates on how the survey sample has changed overall, and then broken down by various demographic characteristics of interest. Regression estimates show no statistically significant differences and provide confidence that significant findings in Table 3.4 are not due to changes in survey responders.

Finally, Column 5 presents estimates for SERP’s impact on total arrest charges. Similar to the suspension outcomes, I use arrests per student (total number of arrest charges divided by total number of students) rather than total arrests to account for varying student population size across the schools¹⁸. It is not clear what impact, if any, SERP would have had on students’ arrest outcomes. There could have be an incapacitation effect where students were less likely to commit certain crimes because they were in school. On the other hand, higher concentrations of

¹⁸ I present arrest charges that occurred during the school year (from September through the following June) in Table 4. Estimates from additional analysis that expanded the time frame to the full 12-month calendar year (rather than the 9-month school year) reveal qualitatively similar results.

students at school could have resulted in more physical altercations (Jacob and Lefgren, 2003). Combined with the national trend towards closer relationships between police departments and schools, and it is possible that SERP could have resulted in overall higher arrest rates (Kim et al., 2010, Nolan, 2011). I find no evidence that SERP had any impact on total arrest rates.

Student-Level Analysis

Research Question 3: Does SERP impact vary by individual risk group? If so, in what ways?

It is possible that SERP has a differential impact on students based on their baseline characteristics. I explore this possibility by sorting students based on their individual risk levels. Appendix E provides the sample size for each combination of school suspension and student risk level. I present all analysis for higher risk students before turning to lower risk students. For each subgroup, I present visual evidence of how the outcomes of interest varied from school year 2010-2011 through 2014-2015 before turning to regression estimates.

High Risk Students

Before presenting results, I follow the school-level analysis plan and formally test for pre-trends (or, the parallel trends assumption) for every outcome by using equation (4). Significant estimates for both high-suspending* SY11 and high-suspending* SY12 would signify a possible pre-trend. If pre-trends were in the same direction as the findings, it would be difficult to tease apart whether the significant findings are due to SERP or pre-existing trends. Appendix Table 3.5 presents the regression results across all outcomes. There are no significant pre-policy interactions, and no patterns of significance that would cause concern for the findings in Tables 3.5 and 3.6 (below).

A. Suspension outcomes

Figure 3.3A presents the differential trend in out-of-school suspensions between high-risk students at high suspending schools and high-risk students at low suspending schools. This graph provides evidence that there were no drastic pre-trends in out-of-school suspensions. In fact, out-of-school suspensions look remarkably stable across the pre-policy years. Furthermore, there appears to be a decrease in out-of-school suspensions after SERP introduction, with larger impacts found in the first full year of policy implementation (SY15). Following what was done with the school-level analysis, I also look at the share of students receiving out-of-school suspensions across time. Figure 3.3B plots the high suspending versus low suspending school difference in the share of students receiving at least one out-of-school suspension across time. This outcome also looks relatively stable in the pre-policy years. Additionally, Figure 3.3B provides preliminary evidence that SERP did not have an impact on the share of students receiving an out-of-school suspension during SY14. However, the figure also suggests that there was a large drop in the share of suspended students during SY15.

Figures 3.3C and 3.3D plot the differential trend in in-school suspensions and the share of students receiving at least one in-school suspension across time. In-school suspensions appear relatively stable in the pre-policy period. The average number of in-school suspensions appears to increase in the post-policy period, though the large standard errors make it difficult to know if these increases are statistically significant. According to Figure 3.3D, SERP did not have any impact on the share of students who received at least one in-school suspension during the year. In fact, this outcome looks remarkably stable throughout the study period.

Table 3.5 displays the regression results for suspension outcomes. These regressions more precisely estimate SERP impact and include school fixed effects. Estimates suggest that

high risk students did not experience a significant impact on both out of school and in-school suspension outcomes during SY14. Larger changes occurred in SY15, the first full year of policy implementation. Out-of-school suspensions dropped by 0.851 suspensions, representing a 49.6% decline from the pre-policy mean. Correspondingly, average out-of-school suspension days dropped by about 2.9 days, a 52.5% decline from the pre-policy mean. Finally, regression analysis suggests that there was a significant shift in the population of students who out-of-school suspensions. There was a 14.5 percentage point decrease (or a -24.7% change from the pre-policy mean) in the share of students who received at least one out-of-school suspension.

At the same time, there is no evidence that SERP caused any significant changes to the number of in-school suspensions, average in-school suspension length, nor the share of students who received an in-school suspension in both post-policy years. Although the estimates are not statistically significant, the point estimates are quite large. For example, the point estimate for in-school suspensions suggest that this outcome may have almost doubled in usage in SY15. Still, the large standard errors mean I cannot rule out the possibility that there were 3.15 more in-school suspensions incidents. But I also cannot rule out the possibility that there were 0.494 fewer in-school suspension incidents.

B. School and arrest outcomes

Table 3.6 displays the regression estimates for other school and arrest outcomes. There is no evidence that SERP significantly impacted high risk students' attendance nor freshmen year academic outcomes. Student reports on classroom order suggest a bit of a rough transition during SY14. Students were significantly less likely to report classroom order during SY14 (-0.233 standard deviation units, a 70% decrease from the pre-policy mean). However, estimates suggest that classrooms quickly adjusted; there were no statistically significant differences in reported

classroom order for SY15. There is also no evidence that SERP significantly changed students' perceptions of school climate. While the point estimates suggest possible improvement in overall school climate for both post policy years, the standard errors are too large to rule out the possibility that school climate deteriorated after SERP introduction.

Column 5 of Table 3.6 presents the regression results for total number of arrest charges. Similar to findings for the out of school suspension outcomes, there are no statistically significant differences in arrest charges for SY14. By SY15 however, there is a significant 0.111 decline in average arrest charges. This change represents a 22.7% decrease from the pre-policy mean.

Lower Risk Students

Following the analysis plan for high risk students, I begin by testing for pre-trends in the outcomes of interest for this subgroup. Appendix Table 3.6 presents the results from this estimation. There is only one statistically significant interaction for the pre-policy years—there is a significant increase in student reports of classroom order during SY12. Combined with the fact that the high-suspending school*SY11 interaction is not significant, I am not worried that there is pre-policy trend in classroom order. Overall, Appendix Table 3.6 provides confidence that significant findings from the regression analysis below (presented in Tables 3.7 and 3.8) are not due to differential trends in the pre-policy period.

A. Suspension outcomes

I begin with visual representations of how out-of-school suspensions and in-school suspensions changed over the study period. Figure 3.4A charts the differential trend in out-of-school suspensions between lower risk students at high suspending schools and lower risk students at low suspending schools. The figure suggests that out-of-school suspensions were

declining slightly in the pre-policy years. It is difficult to tell if the SY14 out-of-school suspension numbers follows the same pattern of decline, or if it is decreasing at a significantly faster rate. Figure 3.4A also provides preliminary evidence that out-of-school suspensions dramatically decreased in SY15. Similarly, Figure 3.4B charts the differential share of students who received at least one out-of-school suspension over time. The share of students suspended remained relatively stable throughout the pre-policy years. Comparable to Figure 3.4A, there is a drop in the share of students suspended in SY14, and an even greater decline in SY15.

Figures 3.4C and 3.4D suggest that despite the changes in out-of-school suspensions, there were few changes in in-school suspensions. Figure 3.4C measures the differential trend in in-school suspensions across time. Similar to previous analysis of in-school suspensions, the standard errors are quite large. Still, in-school suspensions are relatively stable across the pre-policy years, with remarkably little change in the post-policy years. Figure 3.4D, shows the marginal (high-versus-low suspending school) difference in the share of students receiving at least one in-school suspension across time. The share of students receiving an in-school suspension look relatively stable, or slightly increasing, during the pre-policy period. Returning to the regression results in Appendix Table 3.6 shows that although the mean is slightly increasing from SY11 until SY13, this difference is never significantly different from the SY13 mean. Figure 3.4D also suggests that the share of students receiving in-school suspensions is either stable or slightly decreasing in the post-policy years.

Table 3.7 presents the regression results for lower risk students on suspension outcomes. Although estimates for both out-of-school suspensions totals and the share of students suspended suggest declines in these outcomes, neither are statistically significant for SY14. However, the

average length of out-of-school suspensions significantly decreased right after SERP introduction. Average out-of-school suspension length decreased by 0.586 days, a 32.5% decline from the pre-policy mean. Mirroring what was found with both the school-level analysis and the high-risk student subgroup, there are greater impacts during SY15, the first full year of policy implementation. Out-of-school suspension averages significantly declined by 0.412, a 64% decrease from the pre-policy mean. Accordingly, the average length of out-of-school suspensions declined by about one day, representing a 60.5% reduction from the pre-policy mean. Furthermore, there was a significant shift in the share of students receiving at least one out-of-school suspension. There was a 16.8 percentage point drop in the share of suspended students, representing a 53.5% decline from the pre-policy mean.

Closely following the findings for high risk students, there are no significant changes for any of the in-school suspension outcomes. The regressions results for average number of in-school suspensions and average in-school suspension length suggest a possible increase in in-school suspension usage during the post-policy years. However, the standard errors are large and cannot rule out the possibility that in-school suspensions *decreased* in the post-policy years. All in all, Table 3.7 suggests that SERP was successful in decreasing out-of-school suspensions for lower risk students. Additionally, these decreases were not simply a shift from out-of-school suspensions to in-school suspensions.

B. School and arrest outcomes

Table 3.8 reports the regression results for SERP impact on other school and arrest outcomes. There is no evidence that SERP significantly impacted lower risk students' attendance or academic outcomes during SY14. By SY15, there is a significant 0.095 standard deviation improvement in the attendance index. Improved attendance was not paired with any significant

changes for the academic index. Survey analysis shows that overall, students experienced little change in classroom order and school climate in the post-policy years. Point estimates suggest the possibility that school climate deteriorated in SY14 and strengthened (from the pre-policy mean) in SY15, though none of these estimates are statistically significant.

Regression results for arrest charges are shown in the last column (column 5) of Table 3.8. In line with both the school-level analysis and the high-risk subgroup analysis, there is no significant change in arrest charges during SY14. The point estimate suggests a decline, but the standard error prevents me from also ruling out a 0.03 increase in arrest charges in SY14. SERP's impact on arrest charges is decisively clearer for SY15. On average, lower risk students had 0.060 fewer arrest charges in SY15. This significant decrease represents a 63% decline from the pre-policy mean.

Robustness Check

One potential worry of the difference-in-differences method is that the policy could have changed the composition of the treatment and control group. One could imagine a scenario where, if parents knew that some schools are focusing on decreasing out-of-school suspensions and improving socio-emotional skills, they would move their kids to these schools. If there were significant composition shifts where, for example, students with higher test scores shifted their attendance to high suspending schools, it would be impossible to tell whether changes in outcomes are due to SERP, or simply because more high-scoring students enrolled in these schools.

School-Level Test

I test for compositional changes by using the difference-in-differences estimator for each of the baseline covariates. All these characteristics are controlled for in every regression, so

statistically significant differences are not a threat to my research design. However, this test can give a sense of how demographics are or are not changing across time and treatment group. Table 3.9 presents the results for the school-level analysis. The results in column 1 provide evidence that there were no differential changes in gender composition in high suspending schools, compared to low suspending schools, during both post-policy years. In an ideal world, all the coefficients in Table 3.9 would be small and not statistically significant. Non-significance would provide evidence that there were no dramatic pre-post policy changes in student composition at high suspending schools.

The regression results in Table 3.9 suggest that high and low suspending schools experienced relatively similar trends in student composition during the post-policy years. There are no significant differences for the share of students who are male, speak a language other than English at home, are black, are Hispanic, are white, have an IEP, are Emotionally Disturbed, or receive free or reduced lunch. Additionally, there are no significant changes in student population, 8th grade math test scores or share of students arrested in 8th grade. Still, these estimates do suggest a few notable differences. In SY14, freshmen attending high suspending schools had a significant decrease in 8th grade GPA (0.165 decrease, or an 8.8% decline from the pre-policy mean) and 8th grade reading test score (0.104 decrease, or a 19.0% decline from the pre-policy mean). Existing literature find that schools with lower GPAs and test scores are significantly and positively correlated with a higher use of suspensions (Kinsler, 2011; Welch & Payne, 2010; Rocque & Paternoster, 2011; Skiba et al., 2013). This literature suggests that in SY14, high suspending schools had a significant increase in students who were more likely to be

suspended, potentially leading to a negative bias in the data and making it harder to detect any improvements in out-of-school suspensions and other outcomes in SY14.

For SY15, the only significant shift was an increase in students' average 8th grade attendance rates. In SY15, freshmen students at high suspending schools had a 1.2 percentage point increase in their attendance rate. This shift is equivalent to 1.3% increase from the pre-policy mean. Higher attendance rates are negatively correlated with suspension rates (i.e. Stevens et al. 2015), though it is not quite clear how the increased 8th grade attendance should be interpreted here. It is possible that SERP, introduced in SY14, caused attendance rates in SY14 to increase so we could be picking up a policy impact rather than student composition change. I am encouraged that this difference is small, and that there are no statistically significant differences across all the other characteristics. As a sensitivity test, I also run my main analysis (Tables 3.3 through 3.8) with a set of covariates limited to those characteristics that should not be impacted by SERP (the demographics presented in Table 3.9, Columns 1 through 9). The results are qualitatively similar. Given these results, I am confident that my findings are not due to compositional changes across time at high- and low- suspending schools.

Student-Level Test

As with the school-level analysis, one potential worry for the student-level approach is that there may be large changes to these individual risk groups across time, and that any changes in outcomes could be due to these composition changes rather than to the policy change. I use equation 1 to formally test how gender, speaking a language other than English at home, race, having an IEP, free/reduced lunch status, 7th and 8th grade attendance rate, 7th and 8th grade GPA, 8th grade reading and math test scores, and 8th grade arrest history may differentially change over time. For ease of interpretation and to help determine the sign of the potential bias, I also create a

baseline index score by combining all the previously mentioned variables. Items that have been shown to be negatively correlated with suspensions (such as 8th grade attendance rate, GPA, and test scores) are reverse coded. Variables for each index were z-scored by year¹⁹, before being averaged together to form the index. Higher scores on the baseline index represent an increase in suspension risk factors.

Table 3.10 analyzes how the covariates fluctuate over the post policy years, for high risk students only. There were significantly fewer Hispanic students, more students with IEPs, more free or reduced lunch students, and students with lower 8th reading test scores attending high suspending schools in SY14. The baseline index score (Column 15) suggests that, taken together, there is a greater increase in students with higher suspension risk. Correspondingly, regression estimates may be negatively biased, making it more difficult to detect any improvements in out-of-school suspensions and other outcomes in SY14. Looking across the estimates for SY15, the covariates look remarkably similar to the pre-policy trends. The one exception is that there is a significant decrease in the share of the student population who is male, a variable that is associated with higher suspension risk. The baseline index score suggests that, taken together with the rest of the covariates, there is no significant change in student characteristics in SY15. These results give me more confidence that my findings are not due to differential changes in student composition across high- and low- suspending schools in the post-policy years.

Table 3.11 presents the estimates for lower risk students' covariates across time. There were three statistically significant differences in SY14: an increase in the share of students who are male, an increase in the share of students who speak a language other than English at home,

¹⁹ Z-scores are created based on the entire CPS freshmen cohort each year, rather than just the freshmen cohorts used in the analysis.

and a decrease in students' 8th grade GPA. Although share of students who are male and lower GPA is associated with higher suspension risk, speaking a language other than English is associated with lower suspension risk. Looking at the baseline index score (Column 15), we see that overall, these changes were balanced out-- there were no statistically significant differences in SY14. SY15 is even easier to interpret since there were no significant differences across each individual characteristic, nor was there a difference when analyzing the baseline index score. Overall, I am confident that my findings in the main analysis are not due to differential changes in the treatment and comparison groups.

Conclusion

Chicago's Suspension and Expulsion Reduction Plan successfully decreased out-of-school suspension usage without evidence of harm to other student outcomes. By the end of SY15, the first full year of policy implementation, out-of-school suspensions per student declined by 36.5% and there was a 25.2% drop in the share of suspended students compared to the pre-policy mean. Contrary to public fears, these changes were not met with deteriorated school outcomes. There was a significant improvement in the attendance index during SY14 and though estimates remain positive for SY15, it is no longer statistically significant. Additionally, there is no evidence that SERP had a negative impact on freshmen year achievement outcomes. Students reported significantly *less* chaotic classrooms by SY15, and no significant changes in other aspects of school climate. Furthermore, SERP had no impact on arrest charges per student.

Regressions broken down by student-level risk shows that SERP had a slightly different impact on high risk students compared to lower risk students. Both groups experienced significant decreases across all out-of-school suspension outcomes for SY15 and no significant

changes for in-school suspension outcomes. High risk students reported mostly no significant changes across attendance, academic, and school climate outcomes. The one exception is an increase in reported classroom disorder during SY14, though reports returned to pre-policy means by SY15. Lower risk students also reported no significant changes across academic and school climate outcomes. However, there was a significant increase in attendance outcomes in SY15. Finally, both groups had significantly fewer arrests in SY15. These findings suggest that a suspension reduction policy can result in benefits for many students. For lower risk students, having fewer/no suspensions can increase their attendance habits in the short-term, a first step towards increasing school engagement and important outcome by itself.

It is somewhat surprising that despite large declines in out-of-school suspensions, there were no significant improvements in attendance for high risk students. More research should be done on these students facing a double disadvantage (high BSR school and high BSR individual risk) to learn about their barriers for academic success. One hypothesis is that high risk students are already more disengaged in school, and simply reducing suspension sentences is not enough to significantly change their school outcomes. These students may need a more intensive intervention. Alternatively, SERP may have long term benefits for the high-risk group that I was not able to measure due to my data limitations. Future research should examine SERP's impact on longer-term outcomes such as high school graduation and college enrollment.

SERP significantly decreased SY15 arrest charges for both high and lower risk students. Similar to other studies that rely on arrest records, it is difficult to ascertain whether this change was due to fewer arrest-able incidences or if schools/police have shifted their decision-making processes about arrests. Whichever the reason, fewer arrests is an important outcome by itself.

Juvenile arrests are associated with lower levels of school attachment, academic achievement, and high school completion. Furthermore, juvenile arrests are significantly associated with subsequent offending, arrests, and incarceration (Aizer and Doyle, Jr. 2015; Kirk & Sampson 2014; Liberman et al 2014; Mowen and Brent 2016; Sampson and Laub 1997; Wiley et al 2013).

There are a few limitations of this study. One limitation of this study is that teachers and schools have discretion in how/if they want to document student misbehaviors. Some teachers may prefer to never officially log any infractions that may have occurred while others meticulously record most incidents that occur in their classroom. Schools may only record the most serious infractions while choosing to informally deal with most infractions. Alternatively, they may encourage staff to carefully document all misbehaviors. Data for this project comes from CPS misconduct files, which collect all infractions that have been recorded by a teacher/school. While the misconduct files certainly do not capture *all* infractions in a school, it does provide a good introduction to a school's disciplinary environment. Second, much of the previous work on school climate examines teacher reports of school climate. Unfortunately, I was not able to access this data for the current study. Future research should explore how SERP has impacted teachers' views of school climates. It would be interesting to see how similar or different their views are from students' views. Finally, future research should examine how suspension reduction policies impact standardized test scores. It is entirely possible that SERP (or other suspension reduction policies) have an impact on test scores that is completely different from the policy impact on students' grades.

Chapter 4: Talking about Disciplinary Fairness

Introduction

Moral development research suggests that youth are evolving their views on and are more sensitive to issues of fairness during adolescence (Almås et al, 2010; Eisenberg et al, 1995; Eisenberg et al, 2005; Gummerum et al., 2008; Hook & Cook, 1979; Kohlberg, 1993; Sutter, 2007; Turiel, 2005). Additionally, there is also a growing body of literature linking students' perceptions of school rules as fair to positive outcomes such as fewer reports of victimization by both students and teachers, lower levels of misconduct and antisocial behaviors, and higher levels of school engagement and aspirations (Cornell & Huang, 2016; Cornell et al., 2016; Gregory et al., 2012; Gregory et al., 2010; Jia et al., 2016; Welsh, 2000). The current study takes an exploratory look at how students at a Chicago Public High School talked about “fairness” in relation to the school’s disciplinary environment and their experiences throughout the suspension process. My goal is to document the full range of experiences rather than the frequency or commonality of various interactions. In what follows, I summarize the sample and my interviewing techniques and then provide a description of various themes around fairness. In the conclusion, I briefly discuss how future research should further examine how students

conceptualize fairness and how this process can provide insight on what policies and practices schools could adapt to increase their perceptions of fairness.

Data and Methods

School Context

The data come from semi-structured interviews conducted at Smith High School²⁰, a historically higher-suspending public high school in Chicago's South Side. Chicago Public School (CPS) collects data about each school's suspension actions and makes this data publicly available. One commonly used measure of suspensions is the number of suspensions per 100 students (total number of suspensions divided by total number of students and multiplying by 100). This measure allows easy comparison of suspension statistics across schools with varying sizes of student population. During the 2013-2014 and 2014-2015 school years, Smith High School documented 95.93 and 63.61 suspensions per 100 students, respectively. The districtwide average during those same school years was 17.8 and 8.6 suspensions per 100 students, respectively. While overall suspensions have been decreasing throughout the district, and Smith High School is no exception, Smith's suspension numbers remain higher than most other schools in the district. During the 2016-2017 school year (the year of my field research), Smith High School recorded 48.39 suspensions per 100 students—a large decline compared to their suspension numbers from the 2013-2014 or 2014-2015 numbers. Yet, these suspension rates remain high compared to the 2016-2017 districtwide average of 6.7 suspensions per 100 students.

²⁰ School and student names are all pseudonyms.

At the time of the interviews (2016-2017 school year), there were about 280 students enrolled at Smith High School. Most students were low-income (about 95 percent). The student body was about 60 percent black, 28 percent Hispanic, and 10 percent white. Additionally, about four percent of the school was bilingual and 37 percent of the students had Individualized Education Plans. One way Chicago Public Schools evaluates schools' academic environment is through the School Quality Rating Policy score (SQRP). For high schools, the SQRP score is calculated using a weighted scale composed of: the growth in PSAT/SAT test scores, growth of priority groups (i.e. African Americans, Hispanic, etc.) on the SAT, schoolwide average SAT percentile, student attendance, freshman on-track rate, 4-year cohort graduation rate, early college/career credentials, one-year dropout rate, college enrollment, college persistence, 5 Essentials surveys, and data quality. Schools are assigned one of five possible ratings, each with corresponding accountability statuses, based on their SQRP scores: 1+ (good standing), 1 (good standing), 2+ (provisional support), 2 (intensive support), and 3 (intensive support). Provisional support schools may be required to attend additional training and working with the central office on an improvement plan while intensive support schools may be required to take drastic actions such as replacing the principal, being subjected to school turnaround, or school closure. In the year of my field research, Smith High School had a SQRP score of 2.

Recruitment and sample

I worked closely with Smith High School's Attendance Clerk to recruit study participants. Students were considered eligible for interview if their names were recorded into the school disciplinary data for at least one misbehavior in the 2016-2017 school year. The school disciplinary data records all infractions, not only those that resulted in a suspension.

Consequently, some students who were not suspended were also interviewed. The interview pool was created and adjusted on a rolling basis. Every few weeks, Ms. Vasquez would make a new report of the discipline data to see who was recently sent to the dean's office or had incidents logged into the school discipline file. Incident recency was important since I wanted to interview students as close to their suspensions as possible to maximize memory recall of their suspension experience.

Parents and guardians of students on the disciplinary list were contacted and informed about the study. Students under 17 had to bring in waivers signed by their parents while students 18 and over signed their own consent forms. Students could separately agree/disagree to have the interview recorded. I stopped recruiting once I obtained consent forms for 30 students. In total, I reached out to thirty-eight students. No students declined to participate outright, though several students did not remember to bring back consent forms (four) or transferred schools before they were able to be interviewed (three). Three students opted out of a recorded interview. I took detailed notes during these interviews and quickly wrote up a memo after each of these interviews. All recorded interviews were transcribed. Study participants each received a \$20 Visa gift certificate as a token of gratitude for their time.

A total of 30 students were interviewed. 21 students received at least one suspension prior to their interview during the school year. Of the 9 students who were not suspended during the field study year, 7 were suspended in previous school years and were able to speak about those experiences as well as their experiences with other aspects of the disciplinary process. While the remaining two students have never been suspended, they were sent to the office for

various misbehaviors. These students provided insight about their experiences with the overall disciplinary process.

Interview Procedure

Students were asked to describe and make sense of their suspension experiences via a semi-structured, open-ended interview format (see Appendix F for the student interview protocol). Students were asked to talk about their suspension experiences beginning from the most recent infraction incident through when students returned to school. They were also asked about their relationship with various staff members throughout the school, the general school climate, and were prompted to provide suggestions on how they would change the school disciplinary environment.

Interviews were conducted in an empty classroom during school time (usually during gym, lunch, or other non-core class time). I chose to complete the interviews at school, rather than in participants' homes, so that I could easily check in with other students about consent forms, as well as the attendance clerk to get information on more recent misbehavior incidents. Additionally, I hoped (and found to be true) that conducting interviews at schools would be easier for students' own schedules. Interviewed ranged from 20 minutes to 75 minutes, with the mean interview time at around 45 minutes. Since the deans oversee school discipline issues and declined to be interviewed, there are very few interviews that could be triangulated or verified. A few exceptions occurred when disciplinary infractions involved multiple students (i.e. fights) and more than one student decided to participate in the interview study. Consequently, the interviews can only provide data on how the students experienced their disciplinary environment and suspension process.

Coding for “fairness”

Students were asked to describe their suspension process. Sometimes students spoke about the fairness of their suspension on their own accord, while other times they were specifically asked to evaluate its fairness. In cases where students were asked about it directly, students were not given a definition for fairness. Instead, student interpreted fairness however they wished. For the current study, I coded for fairness in multiple ways. In many instances, students simply used the words “fair” or “unfair” in their suspension descriptions. For other students, I coded for instances where students discussed their suspensions in terms of equal treatment, equity, or whether or not students felt their punishments were justified. Altogether, students spoke about fairness in seven broader themes: use of the Student Code of Conduct, infraction-consequence severity match, differing punishments for the same infraction, favoritism or bias, choices, multiple opportunities, and missing work or grades.

Social location/gaining rapport

As an Asian American middle-class woman from the Midwest, I was visibly an outsider to everyone at the school. I use this outsider status to my advantage by feeling free to ask several clarifying questions and asking respondents to explain events in a way they would to their grandmother. I began building rapport with students by explaining that I am a former middle school teacher and that while I was teaching, I was always very curious about students’ suspension experiences. Consequently, I created this study to find out what the suspension experience is like from those who have been through it. I emphasized that they are the experts and are in the best position to talk about what is good (if anything) and what could be improved (if anything) about the suspension process. Before turning on the recorder, I spent the first five

minutes of each interview session reiterating that the interviews are confidential, and that we would not use their name in any interview notes nor on the recording. Instead, I had students select a number that would be attached to their interviews. Most students did not seem worried about the confidentiality aspect of the interviews and a few even asked for their names to be attached to the interviews. These students were eager to talk about their suspensions and asked that their stories be shared directly with the principal²¹.

One worry I had was whether students felt like they had to answer all my questions because I was an adult and that if they did not answer my questions, they would be disobeying an adult. To address this worry, I took care, both before and throughout the interview, to emphasize that students were free to skip any questions they did not wish to answer without any consequences (they would still receive the full amount of compensation and I would not be upset). Additionally, I reiterated that they were able to stop the interview at any time they wished and would still receive their gift card. One student opted to skip a question that he felt was too personal, and no students asked to end the interview early.

Findings

Using the Student Code of Conduct

A few students mentioned the use of the Student Code of Conduct during their suspension process and how this added a sense of fairness to their suspensions. The Student Code of Conduct is the list of rules, behavioral expectations, and consequences for Chicago Public School students. Disciplinary infractions are grouped into six different categories with increasing severity--Group 1 infractions represents inappropriate behaviors and Group 6

²¹ Their stories were not shared with the principal.

represents illegal or seriously disruptive behaviors. Each group of infractions has an assigned list of possible consequences and interventions. In other words, each infraction is not attached to an individual consequence. Instead, deans can choose from a variety of consequences. Students spoke of deans referring to the Student Code of Conduct as one aspect of fairness:

- Natalie: Suspensions can sometimes be fair, depending on situation. But I would say the deans... I respect them and it's because they sometimes look at the code of conduct and they'll go by the rules and I like that.
- Interviewer: Why do you like it?
- Natalie: Sometimes in the Code of Conduct, they sometimes say what a dean should do and the dean does exactly what it says.

For Natalie, looking at the Student Code of Conduct brings fairness to the suspension process. In her opinion, the suspension process can be simplified into finding the infraction in the Student Code of Conduct and finding the consequences listed with each infraction. When the deans follow the Code of Conduct, she views the consequences, even harsh ones like suspensions, as fair since they are merely following the rules previously determined by Chicago Public Schools.

Chris' suspension experience offers a more detailed look at how deans may use the Student Code of Conduct during the suspension experience. Chris was caught smoking marijuana in the bathroom during school hours. When asked to talk about his suspension experience he simply said:

I think it's fair because I coulda got way worse. Cause, once he (Dean Lee) showed me the Student Code of Conduct book, [I saw that the potential consequences were] like 3-5 days suspension, call for expulsion ... and there was a lot of other stuff. But he didn't do none of that. He just gave me the three days. So I was grateful just for the three days cause I coulda got expelled, more days, or anything worse. So I just took the lowest consequence I could get, so yeah, I was happy for the three days.

Dean Lee took the time to pull out the Student Code of Conduct and point out the exact infraction Chris had committed, as well as the corresponding consequences. For Chris, it was

helpful to see that his suspension sentence was explicitly listed in the Student Code of Conduct as well as what other consequences he could have been given. He saw that the school had given him the lightest consequence for his offense, which even made him appreciate his suspension sentence. Although looking through the Student Code of Conduct took more time than simply handing out a suspension, it helped inform his assessment that the suspension was fair. It showed Chris that the deans were following district protocol. What is not clear from the statement is whether Chris would continue to think that the suspension was fair if he was given a five-day suspension instead of his three-day suspension or if he would think the three-day suspension was fair without being presented with the Student Code of Conduct. Still, the combination of these two factors allowed him to view his suspension consequence as justified.

Infraction-Consequence Severity Match

Some students talked about the fairness of their suspensions based on the severity of their infraction. Sometimes students recognized that they were overreacting to a situation and that their actions were serious enough to warrant a suspension. For example, Alexis viewed her suspension as a fair consequence for her infraction. Alexis was involved in a large fight that involved more than five other students:

I feel that the suspension was right, because it was reckless the way we acted. Just, okay, if it was gonna be an argument, it was gonna be an argument, we could've just left it at that, but knowing that we was probably gonna fight, or knowing that it was gonna turn into something else, we could've just waited, and took it after school or something. It looked like we was animals in that lunchroom and I understand the reason why they suspended us. If you could see it on camera, or if you was there to see it, you'd be like ... 'Whoa.'

From Alexis' viewpoint, all the students involved in the fight were out of control. She knew that tension was building amongst all the participants and recognized that they could have just had an argument to air out the tensions. Instead, the situation quickly escalated into a large fight on

school grounds and during school hours. Furthermore, the fight was in the middle of the lunchroom, where there were many witnesses and had the potential to severely disrupt school. Alexis described the fight as “reckless” and described the students as behaving like “animals,” all indications that she saw this as a serious fight. Given the severity of her actions, Alexis saw her suspension as completely justifiable.

Isiah reported a comparable situation, where he felt like his actions warranted a suspension. Isiah also found himself as part of a large fight. He was part of the school’s football team and participated in a fight in the middle of a football game:

Isiah: They (the other team) hit one of our players, they got into a fight so the whole football team gotta fight now.

Interviewer: So the team got suspended together?

Isiah: No, there was too many people to get suspended so they suspended the people that kept fighting. I was one of the people like, "oh you hit me, it's go-time", ain't no pulling me back. Everybody fought, but the people when the coaches grabbed you and said, "hey stay back there" and you stayed ... they gave [those students an] in-school suspensions. But the people that fought and broke off and ran back up (after the coaches told them to stay back) like, "Yeah, what's up? We finna go another round," you getting suspended, because now you're being disobedient. [They already] told us to back off and all that. I think it was pretty decent to me, like the people who keep fighting got suspended, the people that got into a fight but once your coach and stuff told you to stop, you stopped, and so we not gonna [give you an out-of-school suspension] but we gotta do something to you because you got into this fight. So we going to give you in-school suspension, everybody else that want to go round two, you get [an out-of-school] suspension and all that. I guess that was fair enough and all that.

In Isiah’s eyes, the other team started the fight by hitting someone on his team. Still, he recognized that participating in a group fight was against school rules and that students could be suspended for fighting. Additionally, Isiah understood that there were multiple levels of misbehavior in the fight. The first level was just getting in the fight, which Isiah saw as

deserving of a consequence through his comment “we gotta do something.” Then, there were students who did not listen to the coaches when they were asked to stop fighting. In Isiah’s eyes, the second infraction is disobedience. Together with the fighting, this was a more serious set of infractions than just fighting and deserved a harsher consequence than students who stopped fighting when the coaches asked them to stop. All in all, he saw suspension as a fair consequence for his actions since he fought *and* was disobedient.

Other students felt that suspensions were an overly harsh consequence considering their offense. For example, Lisa described the circumstances around her infraction:

I was in the Math class, the teacher wasn't there so this boy was playing music. And it's like, he was right here and I was right here with my phone. The dean walked in, he unplugged it and it looked like I unplugged it from my phone because I was right behind him. So he put me and him out. He was like, "Y'all fixing to get suspended". So I'm like, "Why we fixing to get suspended?". He was like, 'cause I was playing music, but I really wasn't playing music. So he didn't even listen to what I had to say and he just suspended me anyway. [Even if I was playing music, the most he should do is] give me detention because playing music is not really nothing to suspend you for.

Lisa was suspended for playing music in a classroom, even though she reportedly was not the one playing the music and just happened to be nearby the student who was actually playing music²². Regardless of culpability, she believes that suspensions are a serious consequence and should be used only for equally serious infractions. In her opinion, playing music during class is not a suspension-worthy offense. A better-fitting punishment would have been an afterschool detention.

²² After Lisa came back from her one-day suspension, it was revealed that she really was not the person playing music and the suspension was removed from her school records (field notes 11/1/2017).

Similarly, Jason described his suspension as mismatched with his infraction. From Jason's retelling of his suspension, his only offense was talking in class with a friend. The next thing he knew, he was being walked to the dean's office by the teacher:

I mean, they shouldn't, it shouldn't be that serious... He (Dean Lee) just started acting like, uh, I don't know, he ain't show no sympathy. He came in saying about like, y'all going to get some days and bring yall parents in. It's not that serious. Yes, of course I was talking, ain't nothing else to do. Um, I would rather took a detention for talking, take the teacher time, like stay a while afterschool, and figure out what else to do. Do that, that's better than three days suspension.

For Jason, the punishment simply did not fit the crime. In Jason's view, it is difficult for teens to *not* talk in class since there "ain't nothing else to do." Additionally, it did not make sense to Jason that the school would use a harsh punishment, such as a three-day suspension, for such a minor misbehavior. Instead of suspensions, Jason argued, a more reasonable consequence could be a detention or staying afterschool with a teacher. According to Jason, these penalties are more on par with his offense.

Same Infraction, Different Punishments

In some interviews, students spoke about fairness in terms of having consequences equally applied to all students. In other words, these students believed that if multiple students are involved in the same incident, everyone involved should receive the same consequence. Additionally, if other students were involved in a similar incident, they should receive the same consequence. A good illustration of this belief comes from Sean's interview when he described being in a fight with another student but receiving a different suspension sentence:

Sean: I was fightin' with someone (another student). I was fightin' and then I got suspended for three days, he got suspended for like a day. I'm like "Dang!" You feel me? Like, we supposed to get suspended the same [number of] days.

Interviewer: So, you think that you should've got suspended one day or you think he should've been suspended three days?

Sean: We both should've got the same days, however many days we got, we [should have] got the same days.

For Sean, fairness in this situation was not determined by the number of suspension days he was given. He claims not to mind either a one-day or three-day suspension for being in a fight. What he found unjust about his suspension was the fact that the two parties received a different number of suspension days. In Sean's opinion, the two students should have received the exact same consequence.

On the other hand, there were also students who advocated for more varied consequences. In these students' views, consequences should be based on individual students' roles in the incident. Kendra's interview illustrates this alternative view:

Interviewer: Okay, so you were suspended for two days, do you think that was fair?
 Kendra: The two days, it was fair and it was not fair. It was fair because, we caused the fight. But, I feel like it was also unfair because, she had said something rude about Mexican people, and their bodies; where I didn't say nothing about it, you know. I just went up to her because she said something like that. If people were to say something about her race, she's black, so if I was to say something about that, then people would take it offensively. [So,] it is fair because we caused the fight, like we both fought, but it's not fair because she's the one that said the comment, you know, that led to the fight. If she [did not] say that comment, then I would have never went up to her, and we would have never fought.

According to Kendra, she understands that she was suspended for being part of a fight. Part of her sees this as a reasonable consequence for her actions. Yet, a part of her continues to see the suspension as an unfair punishment. From Kendra's perspective, the other student effectively started the fight with a racist comment and she was simply responding to an offensive behavior. In her opinion, there would not have been a fight if the other student did not make a racist comment. Consequently, she sees her own role in the fight as less culpable and the suspension as inequitable.

Martin held a similar complaint when he was suspended for getting into an argument with a classmate.

I don't think [the suspension is] fair because sometimes I didn't even start the argument. The other kid did and they'll be like ...the other kid will get me all worked up and ...They (the other students) know I won't shut up. And they'll (the deans will) just suspend me but sometimes they won't even suspend the other student.

From Martin's perspective, his suspension was unreasonable. For one, he did not start the verbal dispute. According to Martin, the other students already know that he gets easily riled up, yet they are the ones to start arguments with him, suggesting that he is less blameworthy for the fight. Despite this belief, he is sometimes dealt a second source of injustice by being the only person suspended. It is not clear from Martin's statement whether he believes both students should have the same consequence, or if he believes the other student should have been given a harsher consequence than him. Still, it is clear that Martin felt that, given the circumstances, his own suspension was unjust.

Favoritism/Bias

Some students spoke openly about differential treatment and coded these interactions as favoritism or bias. For example, Bianca acknowledged that she can get away with breaking minor rules more easily than her classmates, and often takes advantage of this privilege:

I bring snacks every day to school for me and my friends and in class I get to eat. If somebody else brings snacks, they'd be like, well the teachers like, "Well, put that away." The [other students] always point the finger at me. "Well, Bianca's doing this" and they'd be like, "Well, don't worry about her. I'm talking to you." It'd always be like that...I guess because I'm they favorite. I don't know. Because I'm a good child. I do all my work. I feel like that's the reason why because when I do my work, I do it to the best of my ability. I kid you not, every time I do something I always get 90% or better. There's nothing under that. No C's, no D's. No nothing. I'm always like a A+ worker or whatever. When it comes to me, it's like perfection. I need my stuff to be perfect. That's one of the reasons why and plus a quick worker too. Whenever it's time for us to do something, I go in, I do what I was supposed to do and for the rest of the day, I'd be like ... because I think like, "Okay. If I get this over with now, do it to the best of my ability, I won't have to do it no

more and I'll be done and ain't got to worry about nothing else." It's like one of those and since they (the teachers) see that, I guess they feel well I should be able to do ... I guess it's like a reward, I guess. It's like that, but it's just always the class clowns that like look at you and be like, "Well, how she get to do this and I don't?" Which, sometimes, if I was in their situation, I'd probably say the same thing, but I'm in my situation.

Bianca did not hesitate to acknowledge that teachers treated her differently. They often allow her to eat snacks during class while other students are called out by the teacher and must put their snacks away. She even admitted that if the tables were turned and someone else could eat snacks but she could not, she would similarly question why only some students can break the rules. In Bianca's mind, she was given this special privilege because she was the teacher's "favorite" student. In other words, she saw a positive bias towards her. She goes on to explain that she is probably well-liked by her teacher because she always tried hard in class and did well academically. From Bianca's point of view, the teachers could be allowing her the freedom the snack during class as a reward for being a good student.

While teacher favoritism benefitted Bianca, many students spoke about similar incidents from the other side-- students who got into trouble for committing the same infraction that their classmates participated in. These students talked about being singled out by both teachers and staff infractions. James' case is a good illustration of when a student felt like a teacher was biased against him/her.

James: I start talking (to my classmates next to me), and she (the teacher) tells me shut up or something. So I told her, "Why don't you tell the kids over there to shut up, and don't worry about me, worry about them. You know, there's a whole class over here, you know?" So she goes, "I'm not worried about them, I'm worried about you." And then I'm like, "Why are you so worried about me when you have kids that disrespect you everyday, and you don't nothing?" And then she was like, "Oh, I don't care. If they disrespect me, then they can get out too," or something. So then I didn't worry about her, I was like, all right whatever, I'm not even going to read this book, I'm just going to stand here quiet. Then I stand up the last five

minutes of class, and she's like, "Sit down. I don't care if you don't want to sit down or anything." I'm like, "I don't care, tell all these kids in this classroom to sit down." And then she goes, "You're going to get detention after school," or whatever. So I'm like, "You're going to single me out in front of all these kids?" And then she gave me detention. Favoritism.

Interviewer: You think that it's favoritism?

James: It is. I just told you. Other kids were standing up, and I got a detention out of everybody... I pretty much explained everything to you. It's like most things (about the school's disciplinary environment) are favoritism and stuff. And some kids get left out. And singled out. And I think that's unfair.

According to James, many other students in the classroom were talking to their peers and standing up during the last five minutes of class. Based on his recollections, James even pointed out all the other students who were also standing and talking in the classroom next to him to suggest that he was simply doing the same things as his peers. Yet, the teacher continued to focus solely on his actions and proceeded to assign him a detention. To James, this was a clear example of discrimination. He was not treated in the same manner as his classmates and felt a sense of injustice from this interaction.

Although James reported feeling like bias permeated most disciplinary interactions, many students held more nuanced views of teachers and staff. For example, Mary believed that the deans can be very fair in dealing with some rules and consequences, but more discriminatory about others.

Mary: I think the dean, he like suspend them... Like the day the fight happened he suspend them then he bring them back with their parents either the next day or the day after [to talk about what happened and why].

Interviewer: You think that that's good?

Mary: Yeah, he did that good, but sometimes he'll suspend you like, he'll suspend you if you don't take the school pants. Those school pants be dusty.

Interviewer: Wait, can you explain that more?

Mary: We can't wear like black yoga pants, something comfortable. We can't wear that. Well, most of the girls do, [but] when I wear it, it's such a problem. He sent me home. I walked in the school building one time and

he was like, "Uh-uh, go home." Most of the kids here, they don't get detention for that. Like I saw a girl in leggings yesterday. She didn't get no detention. I asked them like, "Did you get detention for that?" They was like, "No." I was like, "What?!" He looks for me. Every time I try to wear something and sneak into school, like when I wear black leggings or something, he always know. It's like it's always the same thing.

In Mary's opinion, the dean's approach to dealing with fights is equitable--People who fight are suspended and then are also able to talk about the fight when they return. She directly contrasts this situation with how she was treated when she came to school wearing black leggings. Similar to James, Mary felt singled out for wearing black leggings; many other girls wear black leggings and never receive any punishment. She even sought out confirmation for her suspensions by directly asking another student if they received any punishments for wearing the same type of leggings she wore. When she realized that they were not reprimanded, her response, "What?!" accompanied with flailing arms (field notes 11/12/2016), suggested that she did not view this as fair. Mary's comment, "He looks for me," suggests that she feels targeted by the dean, that the dean is always paying special attention to what she is wearing while not paying the same amount of attention to other students. In her opinion, these interactions showed her that not all students are subjected to the same rules about what they can or cannot wear.

Choices

Students also saw their suspensions as fair when they were presented with choices throughout the suspension process. For example, one area that was highlighted by some students was their ability to choose who to talk to when they were in trouble. Perhaps unsurprisingly, students have preferences to talk some adults over others. For example, a student may feel comfortable texting a teacher about a personal family problem, while feeling uncomfortable talking to another teacher about anything non-school related. Natalie's experience provides a

good example of how students linked choices with views of fairness. Natalie got into an argument with a teacher and was asked to leave her class. In her interview, Natalie describes the series of attempts to talk to someone after being asked to leave the classroom:

I ended up going to see a dean and the dean said I was being defiant because I was putting on my coat and he told me not to. He wasn't in a good mood and I saw it in his face and he ended up making me leave the dean office and he walked me downstairs. I sat at the main entrance and then the principal came and I went to the principal office and I was being defiant with him ... I went in the principal office and he was trying to have a discussion with me. I didn't feel like I was wrong because I wasn't in the mood [to think rationally about the situation or to listen to school staff]. But at the end of the day, I admit I was in the wrong because I shouldn't have called out her name (cursed at her) ... I was being defiant towards him and then he ended up calling the officers up. I wasn't talking to the principal at all, I was refusing to talk to him and I was pacing back and forth, so he continued to ask me questions and it came to a breaking point with him where he just didn't put up with it no more. He just called the officers up and then I ended up ... it was two lady officers, and I ended up having a discussion with them. They told me I should have a parent-teacher conference and tell the teacher that she's singling me out. After that, we did have this discussion and it made me [feel] better and even closer [to the teacher]. [...] The principal is fair...He (the principal) gives a lot of options depending on the situation. When I got in trouble, he gave me a chance to talk it out, talk it out with someone else, or call a parent and I chose to talk with someone else.

Natalie's statement shows how students can be upset from being kicked out of the classroom, and how this agitated state can easily carry over to the next adult they interact with. In Natalie's case, she continued to be upset when she met with the dean, and then when she met with the principal. The dean could have easily just punished her for cursing at her teacher, without trying to understand what happened in the classroom. Instead, he decided to let the principal try and talk to her. Although the principal also was not able to determine what happened in the classroom because she was so distressed, he did not resort to an immediate punishment. Instead, he took the time present her with choices about who she would like to talk to. Despite being in a restless state, she realized that she needed to explain what happened to someone, and that her options were quickly dwindling. She forced herself to calm down and ended up telling the two

security guards about the problems she was having with her teacher. While this process was more time consuming than simply handing out a punishment, it treated Natalie with a lot of patience and respect. Consequently, she saw the suspension process as very fair.

Multiple Opportunities

In a similar vein, students spoke about fairness in terms of being given multiple chances. Students believed that if they should be given multiple chances to correct their actions, complete their consequences, and/or gain back their privileges. Sometimes, like in James' case, school officials expected strict adherence to consequences and left no room for discussion. James was suspended after a series of conflicts between him and a teacher. He acknowledged that he was talking in class but emphasized that he felt unfairly singled out by this teacher when the teacher told him to stop talking since other students were also talking and not paying attention. Rather than continuing to participate in classwork, James decided to stand up for the last five minutes of class. As a punishment for this defiant act, the teacher assigned him afterschool detention. James did not go to detention and was given a suspension.

I feel like it's unfair for people to get suspended after not going to detention. I feel like they should have multiple chances, and not just one, to serve a detention. Like what if you have a job application, or anything, or anything could pop up, you know you have to go to the hospital, or something.

In James' opinion, it was unreasonable to suspend students because they missed one detention. There could be several reasons why students miss a detention, yet the school does not allow for students to choose when to serve their detentions. For James, this inflexibility, as well as the extreme consequence for missing detention (suspension), is unfair. He believes that providing students with multiple opportunities to serve detention would be more equitable.

On the other hand, some students experienced disciplinary consequences that gave them a second chance to do better in school and participate in school activities. Kendra's experience provides a good example of such a case. Kendra was suspended for fighting another student in school. Additionally, the deans temporarily withheld her ability to attend senior activities (Smith High School holds a series of events at the end of the school year specifically for seniors such as a special senior luncheon, senior field trip, and prom). Yet, they also provided an avenue for her to gain these senior privileges back—through writing a letter to the deans and having an interview with them.

Interviewer: Are there any consequences you would change?

Kendra: I don't know, the rules are pretty fair. [Here, students] mostly get detention because they're tardy and out of uniform, but over there (at her previous school) it's like for the little things. Over there I know a lot of my friends can't go to prom because they have 13 detentions. Here, if you get in trouble, like how I got in trouble with the fight, I had to write an appeal letter as to why they should let me do my senior activities, like luncheon and prom, and stuff like that, why I should do it. Then, I have to have like an interview, well kind of like an interview with the dean and my principle. I have to say why they should let me go, they're going read my letter, and then I have to explain myself like why they should let me go. Like, they give you a second chance, you know, and over there they don't have that.

Since she recently transferred to Smith from another school, Kendra can compare the disciplinary environment at the two schools. In Kendra's opinion, Smith High School's rules and consequences are fair. She defines fairness in two separate dimensions: the infraction-consequence severity match and the opportunity to regain lost privileges. In Kendra's eyes, Smith High School's consequences are well-matched to the severity of the infractions. For example, she mentions that it makes sense for students to receive detentions for being late to class or out of uniform. She contrasts this with her previous school, where students receive

detentions for “little things,” suggesting that the infractions were, in her opinion, less serious than being late to class. But for her particular case, Kendra views Smith High School as fair because they have given her the chance to attend her senior activities. In order to regain her senior privileges, Kendra is asked to write a letter and go through an interview with the deans to convince them of why they should let her participate in senior activities. Kendra welcomes the opportunity and sees it as “a second chance.”

Missing work and grades

Another aspect of suspension fairness that was brought up in interviews was school work. For some students, making up work was easy. They were presented with a packet of school work when they returned to school and given an adequate amount of time to complete that work for credit. Other students argued that suspensions were not fair because they were missing out on so much school work. Natalie’s interview emphasized this point:

Interviewer: Do you think that the suspension was fair for you?

Natalie: Not really because I didn't learn anything. Kids at school, they have a chance to learn, they gave them extra time to get up on something and... that day I didn't make progress in any class and then I came back [trying to get] missing work that was assigned that day and two days that I could've come to school... I ain't have no homework from a teacher or nothing. That's what I don't like [about] being suspended, I think that we should have an in-school suspension in this school, because when we at home, it's like nothing to do and the teacher wouldn't give us any work because they say they don't have a lesson plan planned out or it was [partner work done in class] or something like that.

For Natalie, suspensions are not fair because it causes her to miss out on school work. In her opinion, suspensions take away her learning opportunities and allows her classmates an academic advantage over her. Her teachers did not provide her with any work during her suspension and offered the following excuses “they don’t have a lesson plan planned out or it

was [partner work done in class].” To Natalie, a more equitable punishment would be an in-school suspension where teachers can still provide students with the classwork for the day. This way, she can serve her punishment without having to miss out on school work.

Discussion

In this current study, I documented the different ways students spoke about fairness, equality, and equity. Students talked about these themes in a variety of contexts throughout the suspension process—from the beginning of the infraction and their interactions with others through coming back to school and trying to make up missing work. There were many students who saw their suspensions as justified – either through the usage of the Student Code of Conduct or because they believed the severity of their actions matched the severity of their consequence. Other students saw their environment as biased and unequal in terms of who was disciplined or not. Some students spoke about fairness in terms of being given multiple choices and chances while still others concentrated on their ability to maintain good grades in school. The breadth of these experiences highlights the multifaceted nature of students’ experiences with the disciplinary process as well as the most memorable, both good and bad, aspects of that experience.

Future research should examine *why* students viewed suspension experiences, or certain aspects of the suspension experience, as particularly fair or unfair. For example, why do students perceive being given multiple opportunities or choices as fairer than rules and consequences with no flexibility? Furthermore, the current literature suggests a correlation between increased sense of rule fairness and higher academic achievement and engagement. Future research should explore *how* students’ perceptions of disciplinary fairness plays a role in their relationships with

school and school staff. It is possible that students may see little or no relationship between fair rules and school engagement. Alternatively, it is possible that when students already have a positive image of/connection to their school, they may see it as a fair place. If future studies find that the relationship between rule fairness and school outcomes is causal, then students' voices about their experiences can shed light on what types of policies or practices are important to their sense of fairness, as well as ways to improve upon the current experiences.

Tables

Table 2.1. Study 1 Summary Statistics

	(1)	(2)	(3)	(4)	(5)	(6)
			Treatment Condition			F-Stat
	Full sample	Black Male	White Male	Black Female	White Female	P-value
Age	36.4567 (12.074)	37.3482 (12.591)	35.9 (11.981)	35.1538 (11.942)	37.5133 (11.776)	0.447
Male	0.5217 (0.500)	0.6071 (0.491)	0.4874 (0.502)	0.5299 (0.501)	0.4643 (0.501)	0.143
White	0.7706 (0.421)	0.7679 (0.424)	0.7417 (0.440)	0.8034 (0.399)	0.7699 (0.423)	0.78
Black	0.0736 (0.261)	0.0446 (0.207)	0.1 (0.301)	0.0855 (0.281)	0.0619 (0.242)	0.432
Teaching Experience	0.1385 (0.346)	0.1696 (0.377)	0.175 (0.382)	0.0855 (0.281)	0.1239 (0.331)	0.145
Location						
Northeast	0.1818 (0.386)	0.1696 (0.377)	0.175 (0.382)	0.2222 (0.418)	0.1593 (0.368)	0.66
Southeast	0.3939 (0.489)	0.3571 (0.481)	0.425 (0.496)	0.3761 (0.487)	0.4159 (0.495)	0.695
Midwest	0.1753 (0.381)	0.1696 (0.377)	0.125 (0.332)	0.1966 (0.399)	0.2124 (0.411)	0.276
West	0.2489 (0.433)	0.3036 (0.462)	0.275 (0.448)	0.2051 (0.406)	0.2124 (0.411)	0.205
Observations	462	112	120	117	113	

Table 2.2. Troublemaker Stereotype Gender Differences

	(1)
Female Mean	-0.116
Male	0.215 (0.094) {0.022}
Observations	460
R-squared	0.034
Covs	X

All specifications include controls for teacher's age, gender, race, teaching experience dummy, and regional Census dummies

Outcomes are standardized by incident with a mean of 0 and standard deviation of 1.

Heteroskedastic robust standard errors. P-values are in curly brackets.

Table 2.3. Troublemaker Stereotype Within Gender Racial Differences

	(1)	(2)
VARIABLES	Male	Female
White Mean	-0.231	-0.494
Black	0.732 (0.121) {0.000}	0.711 (0.123) {0.000}
Observations	231	229
R-squared	0.152	0.195
Covs	X	X

All specifications include controls for teacher's age, gender, race, teaching experience dummy, and regional Census dummies

Outcomes are standardized by incident with a mean of 0 and standard deviation of 1.

Heteroskedastic robust standard errors. P-values are in curly brackets.

Table 2.4. Study 2

Summary Statistics

	(1)	(2)	(3)	(4)	(5)	(6)
			Treatment Condition			F-Stat
	Full sample	Black Male	White Male	Black Female	White Female	p-value
Male	0.3478 (0.4772)	0.3231 (0.4713)	0.375 (0.488)	0.2581 (0.4411)	0.4355 (0.4999)	0.189
White	0.8617 (0.3459)	0.8769 (0.3311)	0.8906 (0.3146)	0.871 (0.338)	0.8065 (0.3983)	0.602
Black	0.0672 (0.2509)	0.0615 (0.2422)	0.0469 (0.213)	0.0645 (0.2477)	0.0968 (0.2981)	0.761
Years Teaching	12.5599 (8.8684)	10.6088 (8.3923)	12.2578 (8.3552)	15.7028 (10.4891)	11.7742 (7.3539)	0.025
Missing: Years Teaching	0.0198 (0.1395)	0.0308 (0.174)	0 (0.000)	0.0484 (0.2163)	0 (0.000)	0.135
Location						
Northeast	0.1621 (0.3692)	0.0923 (0.2917)	0.25 (0.4364)	0.1613 (0.3708)	0.1452 (0.3551)	0.117
Southeast	0.3162 (0.4659)	0.3231 (0.4713)	0.3125 (0.4672)	0.2742 (0.4497)	0.3548 (0.4824)	0.812
Midwest	0.2806 (0.4502)	0.3385 (0.4769)	0.2188 (0.4167)	0.3387 (0.4771)	0.2258 (0.4215)	0.237
West	0.2411 (0.4286)	0.2462 (0.4341)	0.2188 (0.4167)	0.2258 (0.4215)	0.2742 (0.4497)	0.894
Survey Time (in seconds)	589.1423 (440.9958)	634.7385 (714.5765)	545.8281 (211.448)	567.9677 (312.9033)	607.2258 (344.4215)	0.559
Observations	253	65	64	62	62	

Table 2.5. Pooled Incident-Specific Gender

Analysis

	(1)	(2)	(3)	(4)
VARIABLES	Severe	Hindering	Irritating	Discipline
Female Mean	0.017	-0.061	-0.092	-0.061
Male	-0.013 (0.103) {0.898}	0.114 (0.102) {0.265}	0.183 (0.110) {0.097}	0.134 (0.110) {0.226}
Observations	506	506	506	506
R-squared	0.041	0.036	0.059	0.046
Pooled	X	X	X	X

All specifications include controls for teacher's gender, race, years of experience, regional Census dummies, and survey response time.

Outcomes are standardized by incident with a mean of 0 and standard deviation of 1.

Standard errors are clustered by respondent ID. P-values are in curly brackets.

Table 2.6. Incident-Specific Gender

Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Severe		Hindering		Irritating		Discipline	
VARIABLES	First	Second	First	Second	First	Second	First	Second
Female Mean	-0.019	0.054	-0.139	0.018	-0.124	-0.059	-0.103	-0.019
Male	0.091	-0.118	0.283	-0.056	0.258	0.107	0.236	0.032
	(0.123)	(0.125)	(0.125)	(0.120)	(0.124)	(0.125)	(0.125)	(0.127)
	{0.459}	{0.346}	{0.025}	{0.644}	{0.038}	{0.392}	{0.061}	{0.802}
Observations	253	253	253	253	253	253	253	253
R-squared	0.088	0.071	0.066	0.129	0.066	0.104	0.071	0.088

All specifications include controls for teacher's gender, race, years of experience, regional Census dummies, and survey response time.

Outcomes are standardized by incident with a mean of 0 and standard deviation of 1.

Heteroskedastic robust standard errors. P-values are in curly brackets.

Table 2.7. Overall Impression Gender

Analysis

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Impolite	Troublemaker	Angry	Indicative of a Pattern	Dangerous
Female Mean	-0.072	-0.108	-0.094	-0.036	-0.087
Male	0.162 (0.130) {0.213}	0.228 (0.131) {0.082}	0.241 (0.125) {0.055}	0.111 (0.126) {0.378}	0.214 (0.125) {0.089}
Observations	253	253	253	253	253
R-squared	0.036	0.041	0.082	0.059	0.077

All specifications include controls for teacher's gender, race, years of experience, regional Census dummies, and survey response time.

Outcomes are standardized by incident with a mean of 0 and standard deviation of 1. Heteroskedastic robust standard errors. P-values are in curly brackets.

Table 2.8. Racial Differences in Male Incident-Specific Outcomes

(pooled)

	(1)	(2)	(3)	(4)
VARIABLES	Severe	Hindering	Irritating	Discipline
White Mean	-0.138	-0.088	-0.14	-0.106
Black	0.187 (0.143) {0.192}	0.279 (0.125) {0.027}	0.511 (0.148) {0.001}	0.272 (0.149) {0.071}
Observations	258	258	258	258
R-squared	0.110	0.096	0.151	0.146
Pooled	X	X	X	X

All specifications include controls for teacher's gender, race, years of experience, regional Census dummies, and survey response time.

Outcomes are standardized by incident with a mean of 0 and standard deviation of 1. Standard errors are clustered by respondent ID. P-values are in curly brackets.

Table 2.9. Racial Differences in Male Incident-Specific Outcomes (by Incident)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Severity		Hindering		Irritate		Discipline	
VARIABLES	First	Second	First	Second	First	Second	First	Second
White mean	-0.091	-0.186	-0.083	-0.092	-0.158	-0.123	-0.075	-0.137
Black	0.166	0.208	0.404	0.154	0.583	0.440	0.270	0.274
	(0.176)	(0.173)	(0.157)	(0.168)	(0.175)	(0.175)	(0.169)	(0.179)
	{0.348}	{0.232}	{0.011}	{0.360}	{0.001}	{0.013}	{0.112}	{0.128}
Observations	129	129	129	129	129	129	129	129
R-squared	0.222	0.163	0.176	0.242	0.175	0.227	0.230	0.201

All specifications include controls for teacher's gender, race, years of experience, regional Census dummies, and survey response time. Outcomes are standardized by incident with a mean of 0 and standard deviation of 1. Heteroskedastic robust standard errors. P-values are in curly brackets.

Table 2.10. Racial Differences in Male Overall Impression

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Impolite	Troublemaker	Angry	Indicative of a Pattern	Dangerous
White mean	-0.111	-0.119	0.005	-0.132	-0.058
Black	0.419 (0.175) {0.018}	0.432 (0.177) {0.016}	0.295 (0.190) {0.124}	0.424 (0.193) {0.030}	0.256 (0.166) {0.125}
Observations	129	129	129	129	129
R-squared	0.132	0.102	0.131	0.131	0.156

All specifications include controls for teacher's gender, race, years of experience, regional Census dummies, and survey response time.

Outcomes are standardized by incident with a mean of 0 and standard deviation of 1.

Heteroskedastic robust standard errors. P-values are in curly brackets.

Table 2.11. Racial Differences in Female Incident-Specific Outcomes

(pooled)

	(1)	(2)	(3)	(4)
VARIABLES	Severe	Hindering	Irritating	Discipline
White Mean	0.061	-0.071	-0.027	-0.155
Black	-0.059 (0.154) {0.700}	0.102 (0.158) {0.521}	-0.099 (0.173) {0.569}	0.225 (0.174) {0.199}
Observations	248	248	248	248
R-squared	0.075	0.047	0.071	0.036
Pooled	X	X	X	X

All specifications include controls for teacher's gender, race, years of experience, regional Census dummies, and survey response time.

Outcomes are standardized by incident with a mean of 0 and standard deviation of 1.

Standard errors are clustered by respondent ID. P-values are in curly brackets.

Table 2.12. Racial Differences in Female Incident-Specific Outcomes (by Incident)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Severity		Hindering		Irritate		Discipline	
VARIABLES	First	Second	First	Second	First	Second	First	Second
White mean	0.042	0.08	-0.174	0.032	-0.099	0.046	-0.264	-0.046
Black	-0.121	0.002	0.173	0.030	-0.070	-0.128	0.367	0.082
	(0.166)	(0.201)	(0.179)	(0.195)	(0.192)	(0.198)	(0.201)	(0.198)
	{0.468}	{0.993}	{0.336}	{0.876}	{0.716}	{0.521}	{0.070}	{0.678}
Observations	124	124	124	124	124	124	124	124
R-squared	0.190	0.037	0.076	0.084	0.126	0.067	0.049	0.060

All specifications include controls for teacher's gender, race, years of experience, regional Census dummies, and survey response time.

Outcomes are standardized by incident with a mean of 0 and standard deviation of 1.

Heteroskedastic robust standard errors. P-values are in curly brackets.

Table 2.13. Racial Differences in Female Overall Impression

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Impolite	Troublemaker	Angry	Indicative of a Pattern	Dangerous
White mean	-0.072	-0.138	-0.17	0.029	-0.076
Black	0.101 (0.198) {0.611}	0.170 (0.197) {0.390}	0.133 (0.191) {0.488}	-0.114 (0.188) {0.545}	-0.022 (0.180) {0.905}
Observations	124	124	124	124	124
R-squared	0.081	0.051	0.081	0.048	0.094

All specifications include controls for teacher's gender, race, years of experience, regional Census dummies, and survey response time.

Outcomes are standardized by incident with a mean of 0 and standard deviation of 1.

Heteroskedastic robust standard errors. P-values are in curly brackets.

Table 3.1. Out-of-School Suspension per Student by Model Specification

	# OSS/Student (m)		
	(1)	(2)	(3)
Pre-Policy Low Susp Sch Mean		0.189	
Pre-Policy High Susp Sch Mean		1.358	
High Susp School* SY14	-0.185 (0.141)	-0.258 (0.164)	-0.421** (0.126)
High Susp School* SY15	-0.697*** (0.145)	-0.593** (0.166)	-0.495** (0.149)
Observations	145	145	145
R-squared	0.652	0.707	0.753
Covs	No	Yes	Yes
Fixed Effects	No	Year	Year & School
Number of schlid			29

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Table 3.2. School-Level Suspension Outcomes

VARIABLES	(1) # OSS/Stud (m)	(2) OSS Length (m)	(3) Ever OSS (a)	(4) # ISS/Stud (m)	(5) ISS Length (m)	(6) Ever ISS (a)
Pre-Policy High Susp School Mean	1.358	4.356	0.5	1.289	1.814	0.436
High Susp School* SY14	-0.421** (0.126)	-1.629*** (0.410)	-0.091** (0.031)	0.298 (0.265)	0.392 (0.367)	0.009 (0.059)
High Susp School* SY15	-0.495** (0.149)	-2.522*** (0.514)	-0.126** (0.038)	0.978 (0.718)	0.945 (0.829)	0.066 (0.078)
Observations	145	145	145	145	145	145
R-squared	0.406	0.518	0.537	0.351	0.343	0.219
Number of schlid	29	29	29	29	29	29
Covs	All	All	All	All	All	All

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Table 3.3. Racial Breakdown of Out-of-School Suspensions

VARIABLES	(1) Share black male ≥1 OSS	(2) Share black female ≥1 OSS	(3) Share Hispanic male ≥1 OSS	(4) Share Hispanic female ≥1 OSS	(5) Share white male ≥1 OSS
Pre-Policy High Susp School Mean	0.57	0.467	0.169	0.151	0.05
High Susp School* SY14	-0.112 (0.061)	0.019 (0.062)	-0.165 (0.101)	-0.001 (0.088)	0.058 (0.049)
High Susp School* SY15	-0.109 (0.063)	-0.011 (0.058)	0.046 (0.053)	-0.106 (0.105)	0.088 (0.054)
Observations	145	145	145	145	145
R-squared	0.344	0.361	0.188	0.172	0.154
Number of schlid	29	29	29	29	29
Covs	All	All	All	All	All

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Table 3.3. Racial Breakdown of Out-of-School Suspensions (continued)

VARIABLES	(6) Share white female >=1 OSS	(7) Black/White Male Gap	(8) Black/White Female Gap	(9) Hisp/White Male Gap	(10) Hisp/White Female Gap
Pre-Policy High Susp School Mean	0.059	0.52	0.408	0.119	0.092
High Susp School* SY14	-0.032 (0.070)	-0.171* (0.073)	0.051 (0.094)	-0.224 (0.119)	0.031 (0.104)
High Susp School* SY15	0.105 (0.067)	-0.198* (0.075)	-0.116 (0.091)	-0.043 (0.056)	-0.211 (0.125)
Observations	145	145	145	145	145
R-squared	0.172	0.210	0.196	0.136	0.207
Number of schlid	29	29	29	29	29
Covs	All	All	All	All	All

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Table 3.4. School and Arrest Outcomes

VARIABLES	(1) Attendance Index	(2) Academic Index	(3) Classroom Order	(4) School Climate Index	(5) Total Arrest Charges Per Stud
Pre-Policy High Susp School Mean	-0.267	-0.087	-0.384	-0.256	0.367
High Susp School* SY14	0.162** (0.058)	0.036 (0.089)	0.030 (0.113)	0.118 (0.076)	-0.002 (0.033)
High Susp School* SY15	0.105 (0.062)	-0.047 (0.085)	0.407** (0.135)	0.124 (0.091)	-0.039 (0.042)
Observations	145	145	145	145	145
R-squared	0.365	0.395	0.333	0.188	0.308
Number of schlid	29	29	29	29	29
Covs	All	All	All	All	All

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Table 3.5. Suspension Outcomes, High Risk Students Only

VARIABLES	(1) # OSS (m)	(2) OSS Length (m)	(3) Ever OSS (a)	(4) # ISS (m)	(5) ISS Length (m)	(6) Ever ISS (a)
Pre-Policy High Susp School Mean	1.714	5.506	0.588	1.403	2.12	0.528
High Susp School* SY14	-0.236 (0.215)	-1.155 (0.593)	0.018 (0.048)	0.627 (0.375)	0.792 (0.510)	0.006 (0.077)
High Susp School* SY15	-0.851* (0.329)	-2.888*** (0.720)	-0.145* (0.068)	1.328 (0.911)	1.489 (1.032)	-0.001 (0.104)
Observations	6,202	6,202	6,202	6,202	6,202	6,202
R-squared	0.103	0.111	0.074	0.074	0.056	0.021
Number of schlid	29	29	29	29	29	29
Model	Hi Studs Only	Hi Studs Only	Hi Studs Only	Hi Studs Only	Hi Studs Only	Hi Studs Only

Robust standard errors in
parentheses

*** p<0.001, ** p<0.01, * p<0.05

Table 3.6. School and Arrest Outcomes, High Risk Students Only

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Attendance Index	Academic Index	Classroom Order	School Climate Index	Total Arrest Charges
Pre-Policy High Susp School Mean	-0.368	-0.367	-0.329	-0.249	0.488
High Susp School* SY14	0.013 (0.098)	0.007 (0.114)	-0.233** (0.080)	0.127 (0.079)	-0.033 (0.065)
High Susp School* SY15	0.050 (0.072)	-0.075 (0.095)	0.130 (0.157)	0.053 (0.075)	-0.111* (0.042)
Observations	6,202	6,202	3,307	3,307	6,202
R-squared	0.301	0.229	0.023	0.028	0.239
Number of schlid	29	29	29	29	29
Model	Hi Studs Only	Hi Studs Only	Hi Studs Only	Hi Studs Only	Hi Studs Only
Covs	All	All	All	All	All

Robust standard errors in
parentheses

*** p<0.001, ** p<0.01, * p<0.05

Table 3.7. Suspension Outcomes, Lower Risk Students Only

VARIABLES	(1) # OSS (m)	(2) OSS Length (m)	(3) Ever OSS (a)	(4) # ISS (m)	(5) ISS Length (m)	(6) Ever ISS (a)
Pre-Policy High Susp School Mean	0.644	1.802	0.314	0.525	0.723	0.267
High Susp School* SY14	-0.231 (0.120)	-0.586** (0.209)	-0.079 (0.054)	0.191 (0.137)	0.209 (0.158)	0.014 (0.047)
High Susp School* SY15	-0.412*** (0.063)	-1.091*** (0.253)	- 0.168** *	0.181 (0.279)	0.166 (0.287)	-0.003 (0.055)
Observations	19,736	19,736	19,736	19,736	19,736	19,736
R-squared	0.062	0.047	0.060	0.040	0.041	0.035
Number of schlid	29	29	29	29	29	29
Model	Not Hi Only	Not Hi Only	Not Hi Only	Not Hi Only	Not Hi Only	Not Hi Only

Robust standard errors in
parentheses

*** p<0.001, ** p<0.01, * p<0.05

Table 3.8. School and Arrest Outcomes, Lower Risk Students Only

VARIABLES	(1) Attendance Index	(2) Academic Index	(3) Classroom Order	(4) School Climate Index	(5) Total Arrest Charges
Pre-Policy High Susp School Mean	0.064	0.267	-0.203	-0.217	0.095
High Susp School* SY14	0.083 (0.043)	0.123 (0.087)	-0.113 (0.140)	-0.007 (0.070)	-0.027 (0.015)
High Susp School* SY15	0.095* (0.037)	0.136 (0.138)	0.150 (0.127)	0.049 (0.065)	-0.060** (0.019)
Observations	19,736	19,736	15,474	15,474	19,736
R-squared	0.314	0.302	0.023	0.018	0.048
Number of schlid	29	29	29	29	29
Model	Not Hi Only	Not Hi Only	Not Hi Only	Not Hi Only	Not Hi Only
Covs	All	All	All	All	All

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Table 3.9. School-Level Characteristics

Over Time

VARIABLES	(1) Male	(2) Oth Lang @ Home	(3) Black	(4) Hispanic	(5) White	(6) IEP	(7) Emotionally Disturbed	(8) Free/ Reduced Lunch
Pre-Policy High Susp Sch Mean	0.531	0.074	0.879	0.086	0.026	0.238	0.022	0.963
High Susp School* SY14	0.029 (0.023)	0.012 (0.015)	-0.008 (0.012)	-0.001 (0.014)	-0.002 (0.004)	0.024 (0.022)	0.013 (0.009)	0.000 (0.019)
High Susp School* SY15	-0.022 (0.018)	0.012 (0.017)	-0.018 (0.016)	0.006 (0.014)	-0.003 (0.009)	0.032 (0.022)	0.004 (0.008)	0.013 (0.018)
Observations	145	145	145	145	145	145	145	145
R-squared	0.064	0.025	0.122	0.060	0.011	0.140	0.059	0.041
Number of schlid	29	29	29	29	29	29	29	29

Robust standard errors in
parentheses*** p<0.001, ** p<0.01, *
p<0.05

**Table 3.9. School-Level Characteristics Over Time
(continued)**

VARIABLES	(9) Student Population	(10) GPA (8th)	(11) Reading test (8th)	(12) Math test (8th)	(13) Attendance rate (8th)	(14) Any arrest charge (8th)
Pre-Policy High Susp Sch Mean	160.952	1.881	-0.547	-0.571	0.911	0.113
High Susp School* SY14	-2.448 (12.296)	-0.165* (0.071)	-0.104* (0.050)	0.010 (0.043)	-0.009 (0.006)	0.023 (0.014)
High Susp School* SY15	-14.676 (20.968)	-0.089 (0.077)	0.012 (0.046)	-0.006 (0.049)	0.012* (0.005)	0.004 (0.018)
Observations	145	145	145	145	145	145
R-squared	0.220	0.108	0.213	0.139	0.403	0.088
Number of schlid	29	29	29	29	29	29

Robust standard errors in
parentheses

*** p<0.001, ** p<0.01, * p<0.05

Table 3.10. Compositional Changes in "High Risk"

Students

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Male	Oth Lang @ Home	Black	Hispanic	IEP	Free/ Reduced Lunch	8th Grade GPA	7th Grade GPA
Pre-Policy High Susp Sch Mean	0.604	0.017	0.956	0.033	0.265	0.985	1.646	1.431
High Susp School* SY14	-0.054 (0.039)	-0.004 (0.030)	0.037 (0.019)	-0.050* (0.022)	0.115** *	0.019* (0.009)	-0.100 (0.049)	-0.076 (0.041)
High Susp School* SY15	-0.084* (0.036)	-0.005 (0.027)	-0.014 (0.025)	-0.022 (0.031)	0.059 (0.029)	0.018 (0.012)	-0.039 (0.077)	-0.042 (0.041)
Observations	6,202	6,202	6,202	6,202	6,202	6,202	6,202	6,202
R-squared	0.002	0.000	0.003	0.004	0.004	0.002	0.003	0.008
Number of schlid	29	29	29	29	29	29	29	29
Model	Hi Studs Only	Hi Studs Only	Hi Studs Only	Hi Studs Only	Hi Studs Only	Hi Studs Only	Hi Studs Only	Hi Studs Only

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Table 3.10. Compositional Changes in "High Risk" Students (continued)

VARIABLES	(9) 8th Grade Reading Test Score	(10) 8th Grade Math Test Score	(11) 8th Grade Attendance Rate	(12) 7th Grade Attendance Rate	(13) 8th Grade Misdemeanor	(14) 8th Grade Felony	(15) Baseline Index
Pre-Policy High Susp Sch Mean	-0.742	-0.744	0.903	0.905	0.177	0.068	0.619
High Susp School* SY14	-0.138* (0.050)	-0.082 (0.053)	0.004 (0.008)	0.005 (0.006)	0.058 (0.036)	0.029 (0.016)	0.074*** (0.019)
High Susp School* SY15	-0.005 (0.061)	0.067 (0.061)	0.008 (0.007)	0.002 (0.006)	0.011 (0.032)	-0.003 (0.016)	-0.000 (0.028)
Observations	6,202	6,202	6,202	6,202	6,202	6,202	6,202
R-squared	0.004	0.002	0.009	0.007	0.002	0.002	0.011
Number of schlid	29	29	29	29	29	29	29
Model	Hi Studs Only	Hi Studs Only	Hi Studs Only	Hi Studs Only	Hi Studs Only	Hi Studs Only	Hi Studs Only

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, *

p<0.05

Table 3.11. Compositional Changes in "Lower Risk"

Students

VARIABLES	(1) Male	(2) Oth Lang @ Home	(3) Black	(4) Hispanic	(5) IEP	(6) Free/ Reduced Lunch	(7) 8th Grade GPA	(8) 7th Grade GPA
Pre-Policy High Susp Sch Mean	0.393	0.246	0.614	0.269	0.146	0.907	2.658	2.4
High Susp School* SY14	0.051** (0.017)	0.038** (0.012)	-0.028 (0.016)	0.020 (0.014)	-0.027 (0.018)	0.012 (0.018)	-0.118* (0.047)	-0.085 (0.066)
High Susp School* SY15	0.001 (0.029)	0.010 (0.014)	0.001 (0.010)	0.002 (0.022)	0.015 (0.045)	0.011 (0.012)	-0.046 (0.053)	-0.079 (0.078)
Observations	19,736	19,736	19,736	19,736	19,736	19,736	19,736	19,736
R-squared	0.000	0.001	0.000	0.000	0.002	0.001	0.003	0.011
Number of schlid	29	29	29	29	29	29	29	29
Model	Not Hi Only	Not Hi Only	Not Hi Only	Not Hi Only	Not Hi Only	Not Hi Only	Not Hi Only	Not Hi Only
Covs	All	All	All	All	All	All	All	All

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, *

p<0.05

**Table 3.11. Compositional Changes in "Lower Risk" Students
(continued)**

VARIABLES	(9) 8th Grade Reading Test Score	(10) 8th Grade Math Test Score	(11) 8th Grade Attendance Rate	(12) 7th Grade Attendance Rate	(13) 8th Grade Misdemeanor	(14) 8th Grade Felony	(15) Baseline Index
Pre-Policy High Susp Sch Mean	0.028	-0.036	0.958	0.959	0.012	0.004	-0.005
High Susp School* SY14	-0.062 (0.051)	0.014 (0.055)	0.001 (0.003)	0.000 (0.002)	0.012 (0.010)	-0.000 (0.003)	0.013 (0.019)
High Susp School* SY15	0.007 (0.096)	0.014 (0.080)	-0.000 (0.007)	0.001 (0.002)	-0.005 (0.006)	0.006 (0.009)	0.014 (0.026)
Observations	19,736	19,736	19,736	19,736	19,736	19,736	19,736
R-squared	0.003	0.003	0.019	0.005	0.001	0.000	0.008
Number of schlid	29	29	29	29	29	29	29
Model	Not Hi Only	Not Hi Only	Not Hi Only	Not Hi Only	Not Hi Only	Not Hi Only	Not Hi Only
Covs	All	All	All	All	All	All	All

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Table 4.1. Study Participant Details				
Name	Gender	Year	Suspended this Year?	Previous Suspensions?
Lisa	F	sophomore	Yes	No
Vanessa	F	senior	No	Yes
Sarah	F	freshman	Yes	No
Isaac	M	freshmen	Yes	Yes
Sean	M	junior	Yes	Yes
Isiah	M	senior	Yes	Yes
Louis	M	senior	Yes	Yes
Ryan	M	senior	Yes	Yes
Trey	M	senior	Yes	No
Elaine	F	freshman	No	No
Nicole	F	sophomore	Yes	Yes
Bianca	F	senior	No	Yes
Mary	F	sophomore	No	No
Natalie	F	sophomore	Yes	Yes
James	M	senior	Yes	Yes
Martin	M	Junior	Yes	Yes
Kendra	F	senior	Yes	Yes
Alyssa	F	sophomore	No	Yes
Amanda	F	freshman	No	Yes
Janelle	F	sophomore	Yes	Yes
Matthew	M	senior	No	Yes
Michelle	F	junior	Yes	Yes
Jason	M	sophomore	Yes	No
Alexis	F	freshmen	Yes	Yes
Melissa	F	sophomore	No	Yes
Nathan	M	junior	No	Yes
Chris	M	freshmen	Yes	No
Jamie	F	freshman	Yes	No
Michael	M	sophomore	Yes	Yes
Andre	M	freshman	Yes	Yes
* I consciously chose not to directly ask students about their race.				
Since suspensions/getting into trouble is already often seen through racial or				
gendered lens, I did not want to start interviews with students thinking about				
stereotype threat. That said, many students naturally brought up race				
throughout the interview.				

Figures

Figure 1

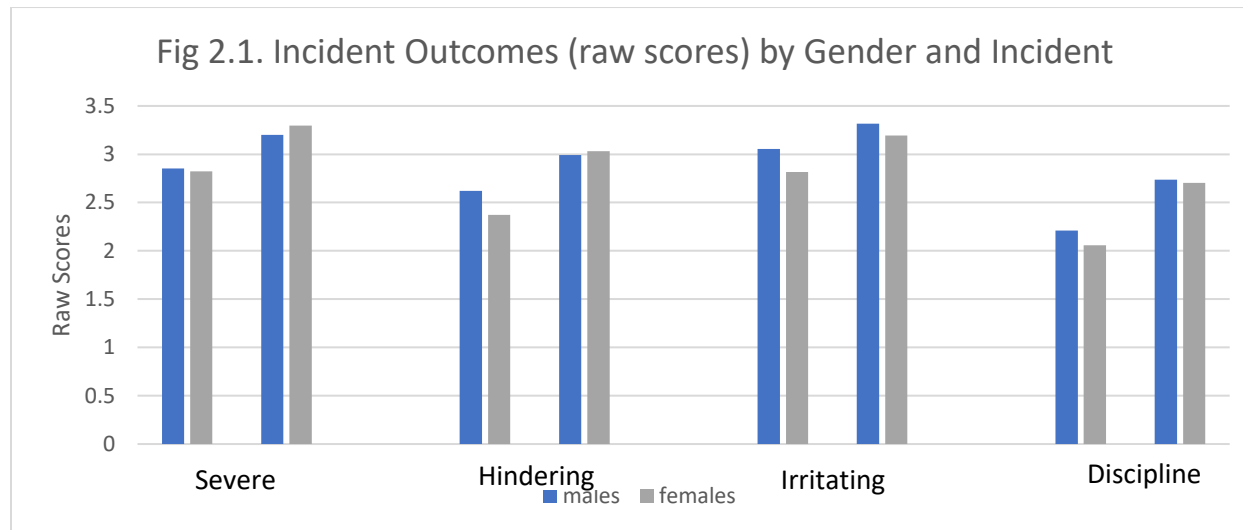


Figure 2

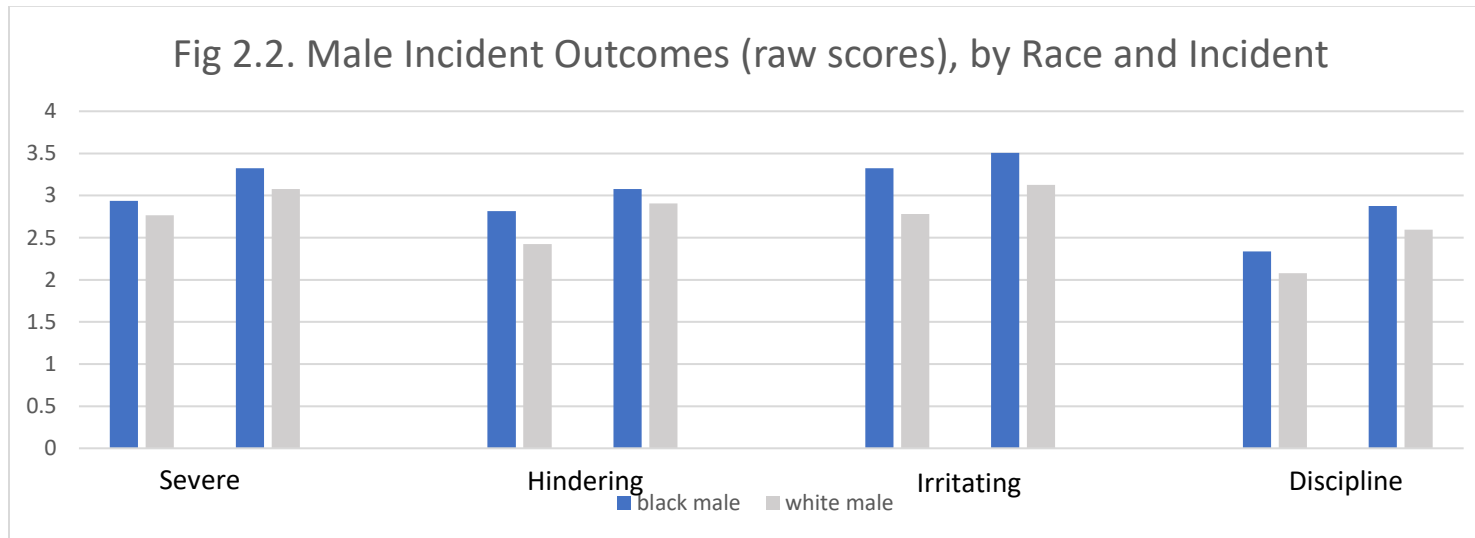


Figure 3

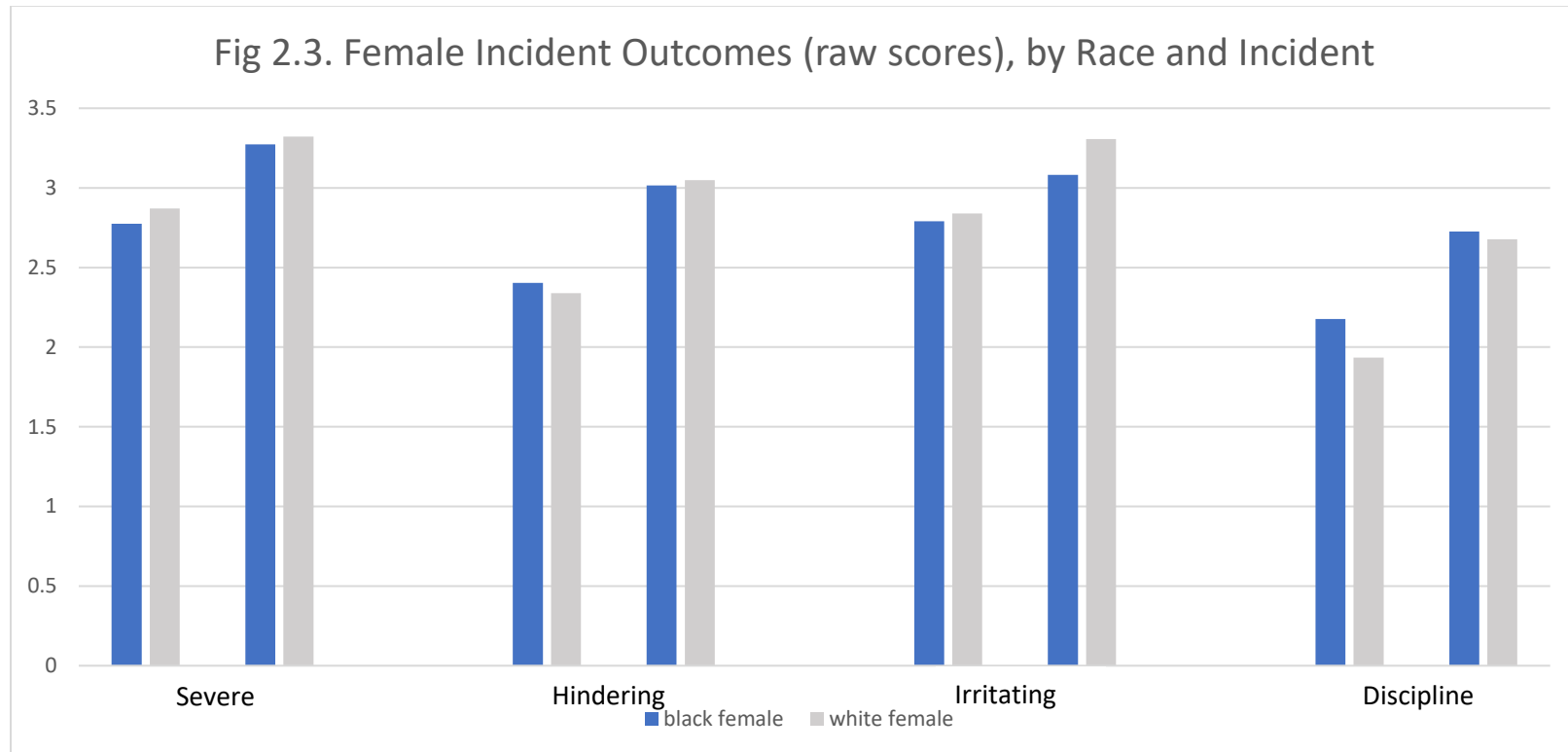


Figure 4

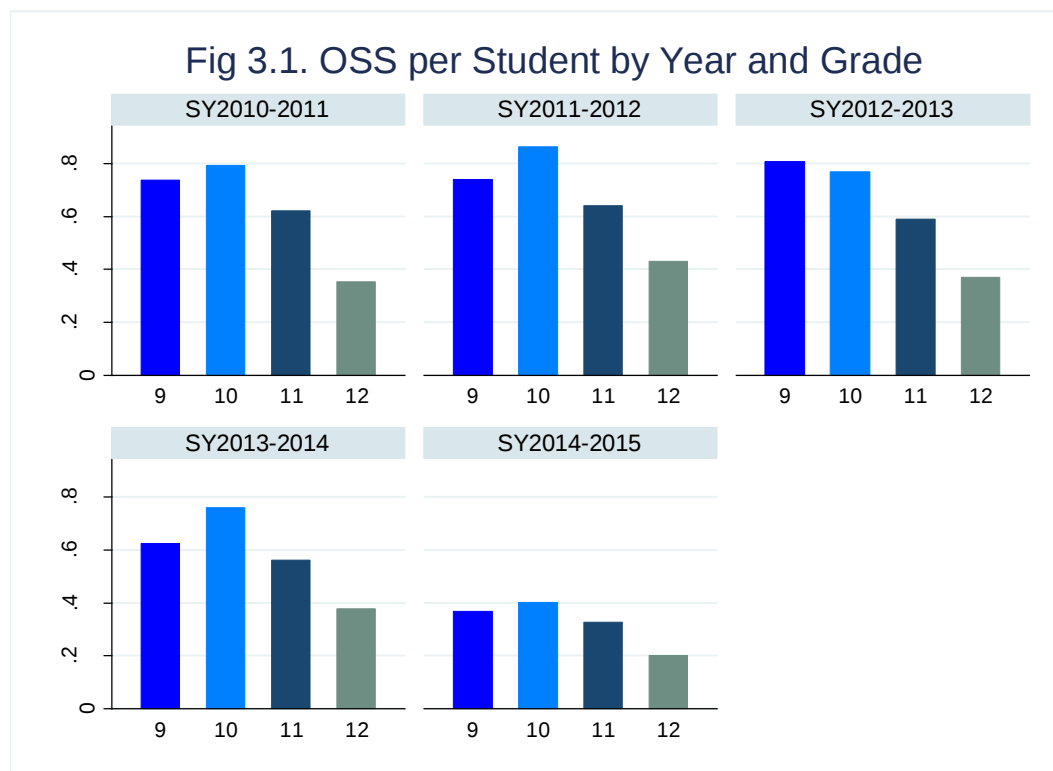


Figure 5

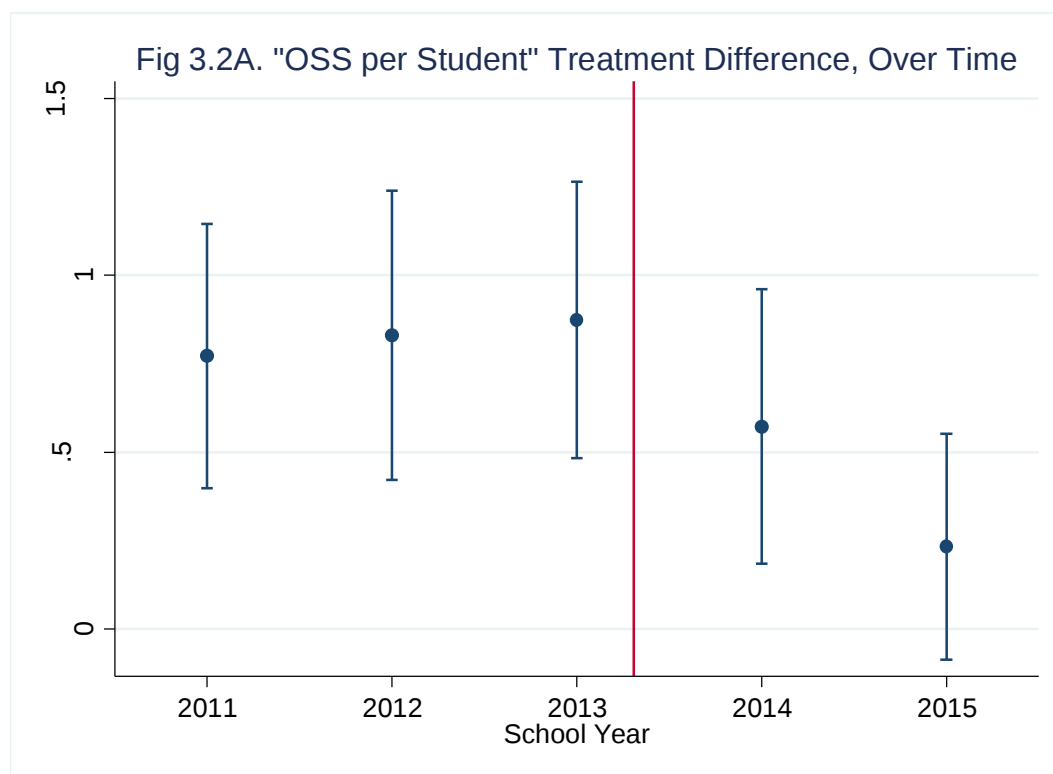


Figure 6

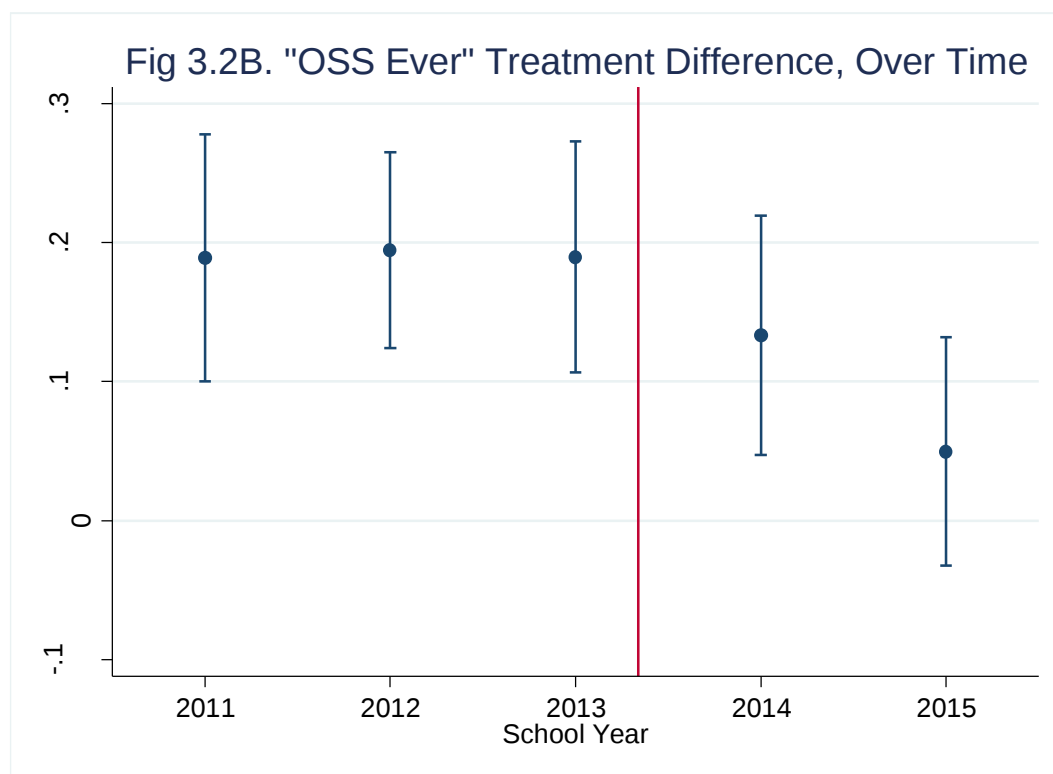


Figure 7

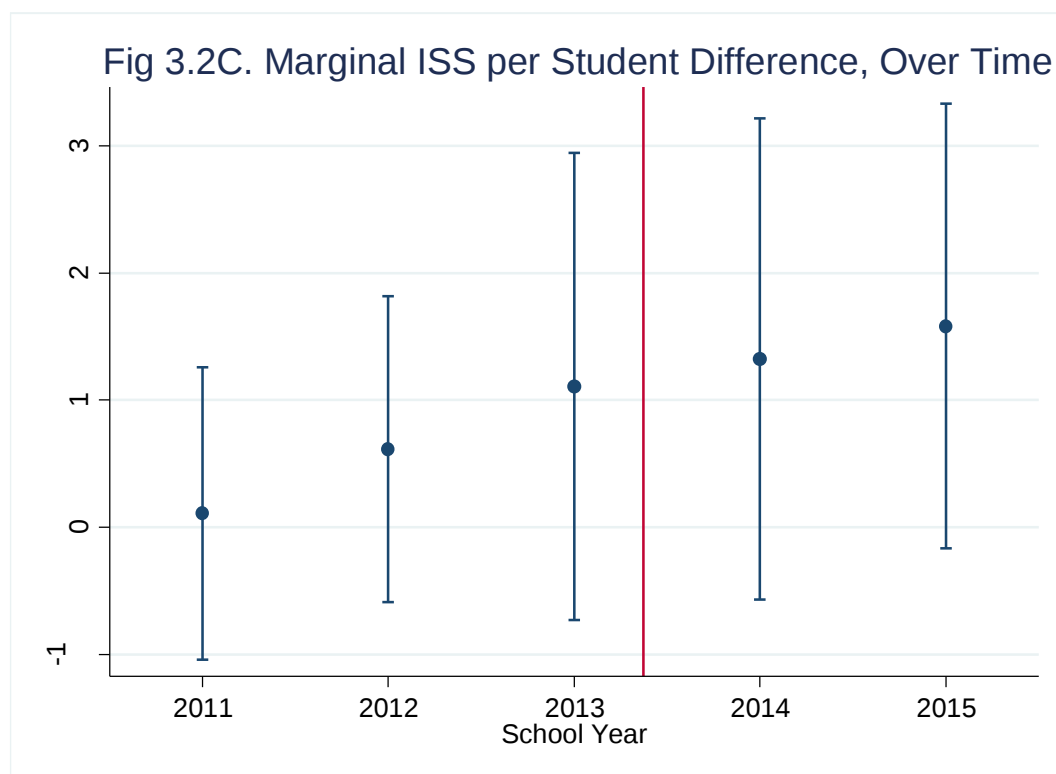


Figure 8

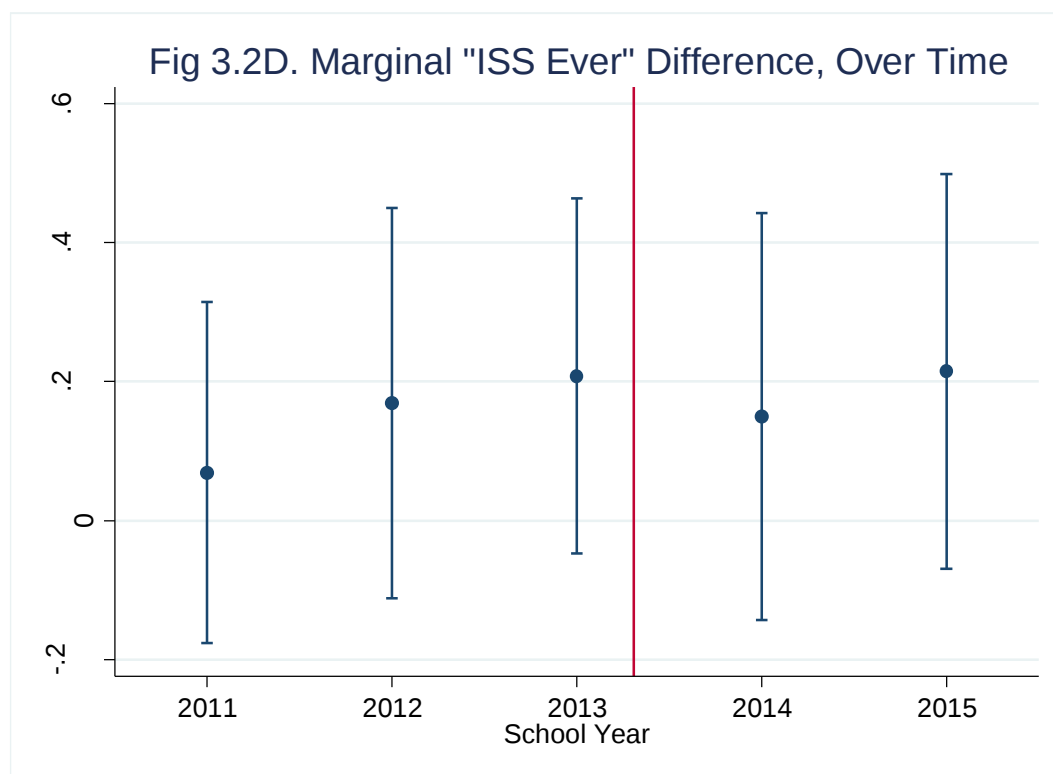


Figure 9

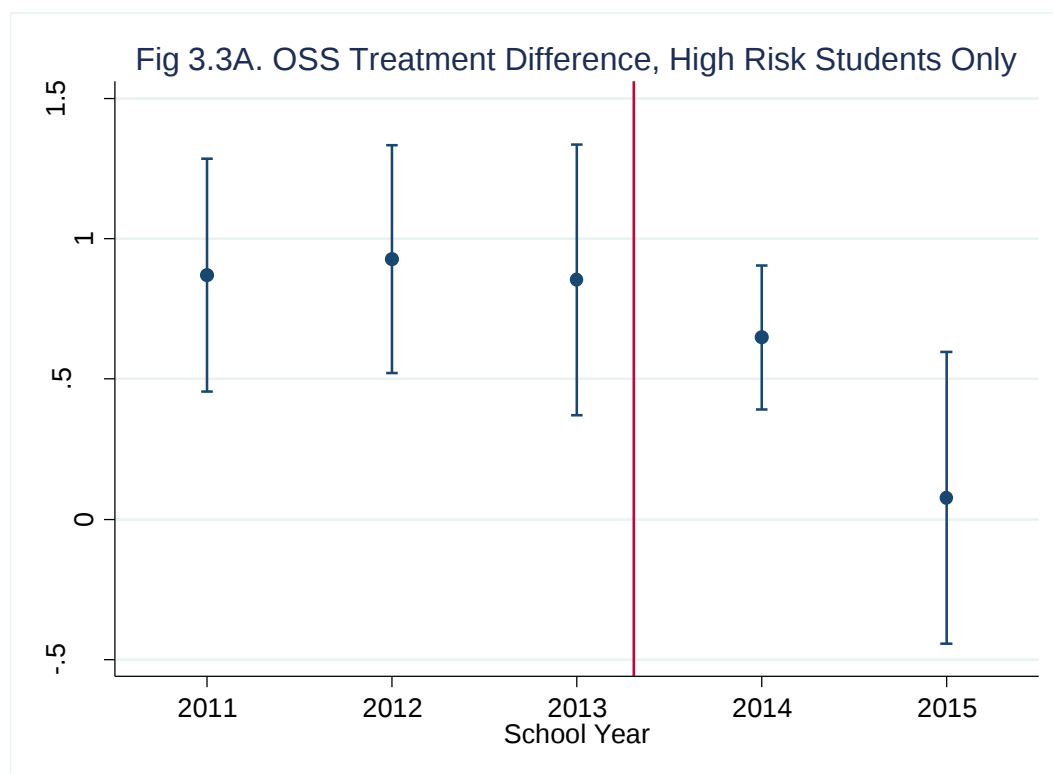


Figure 10

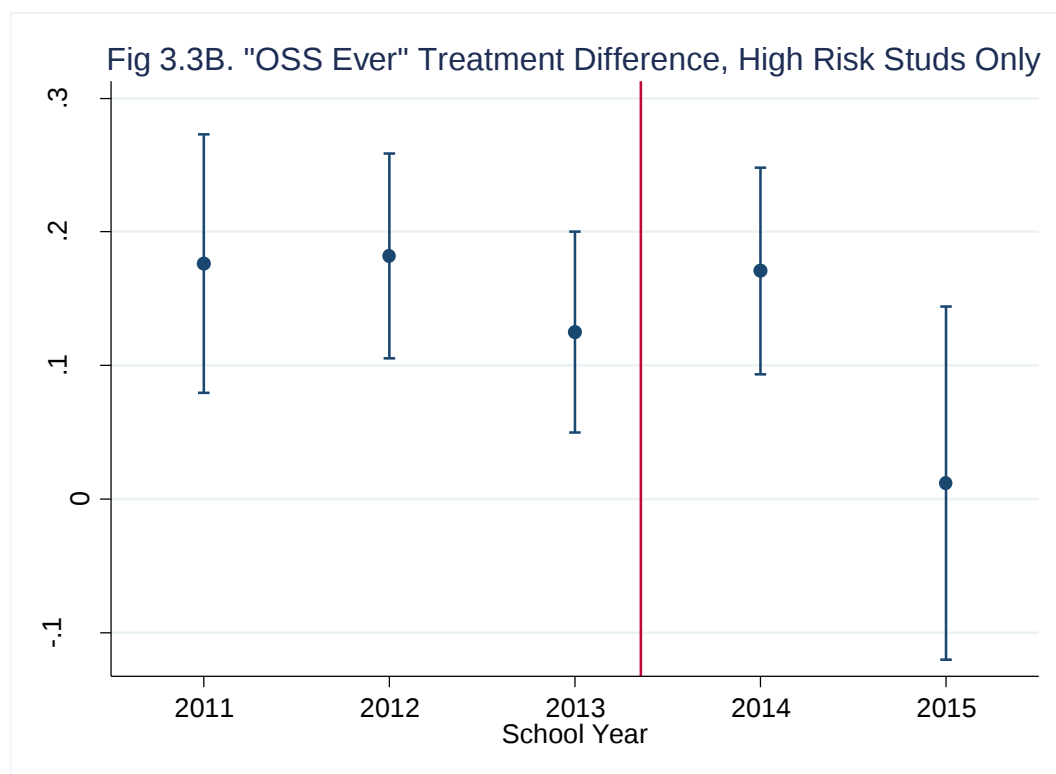


Figure 11

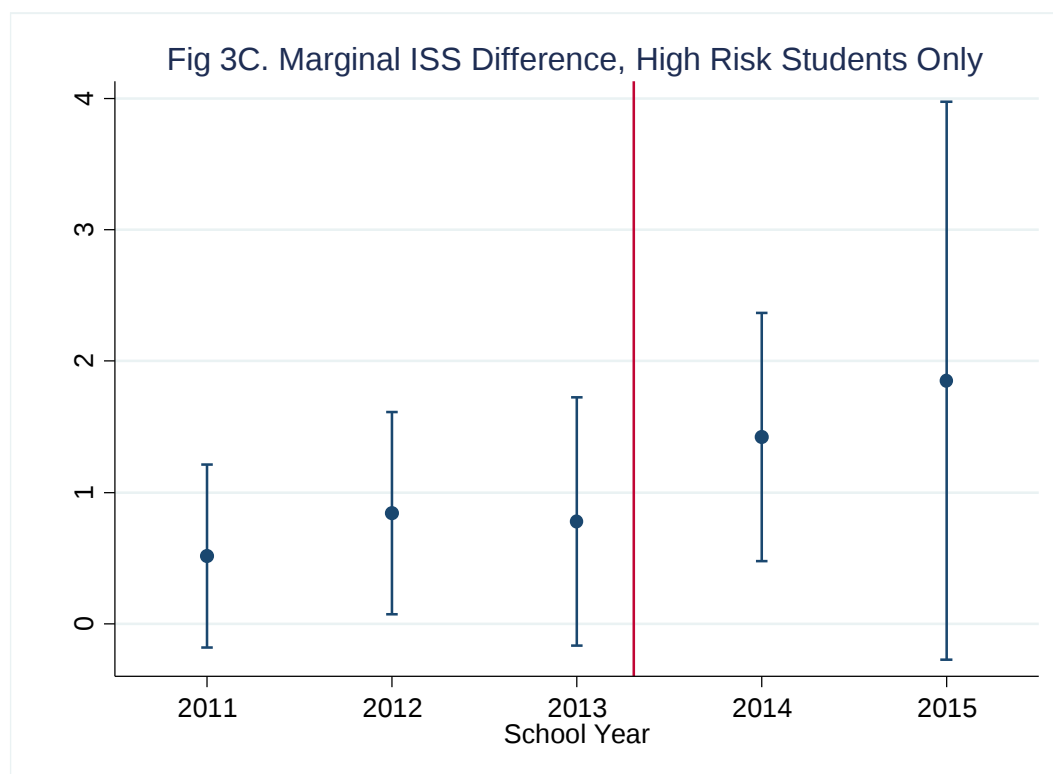


Figure 12

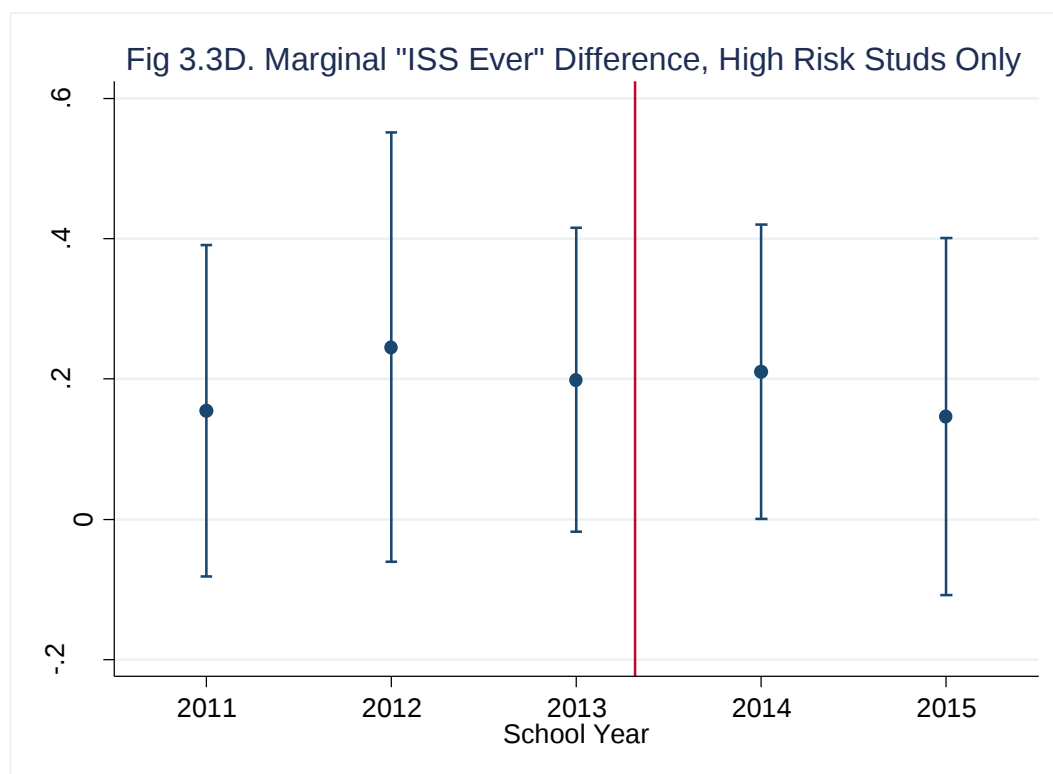


Figure 13

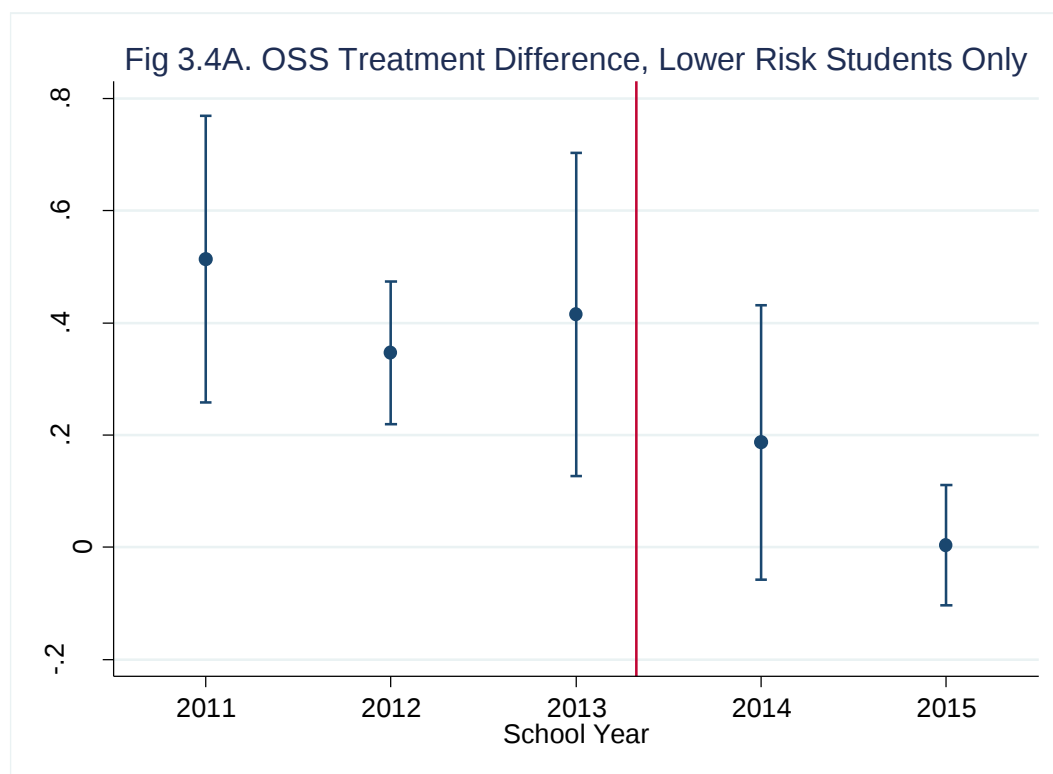


Figure 14

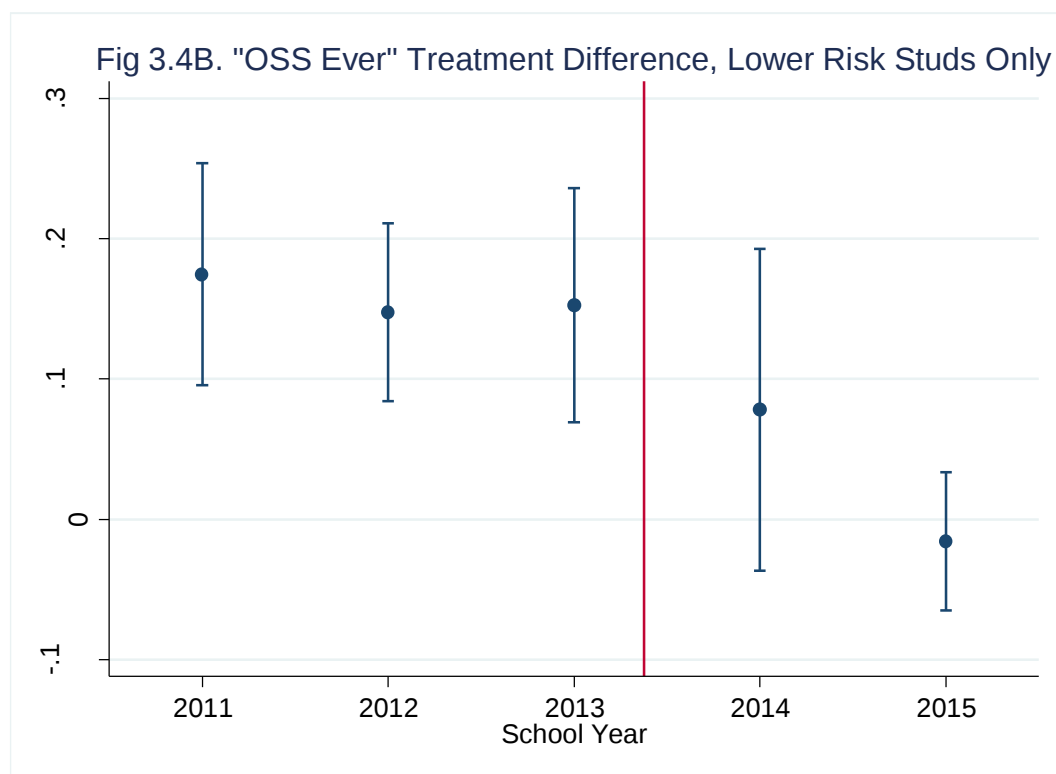


Figure 15

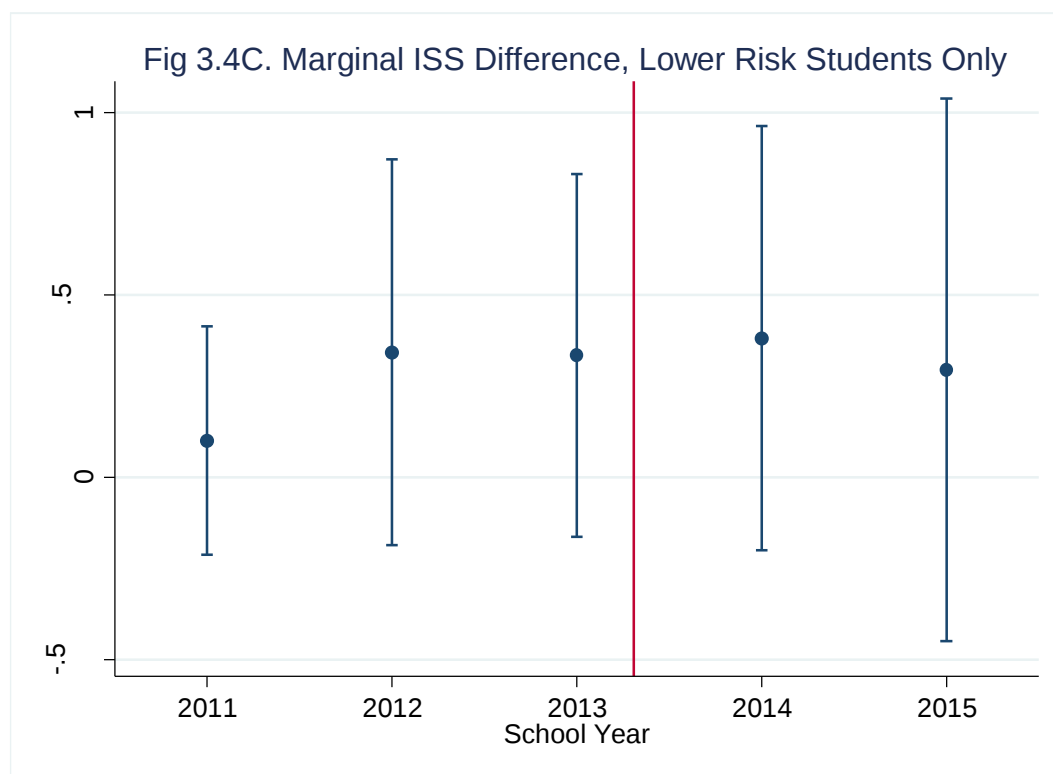
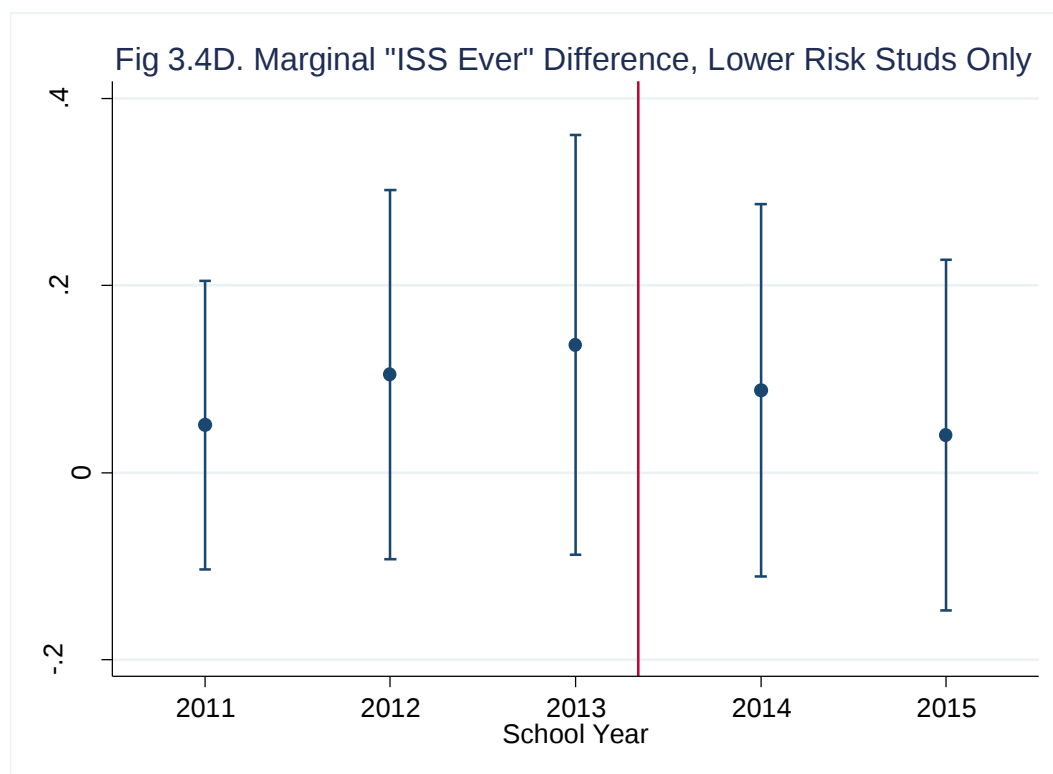


Figure 16



References

- Advancement Project & The Civil Rights Project (2000). Opportunities suspended: The devastating consequences of zero tolerance and school discipline policies. Cambridge, MA: Harvard Civil Rights Project.
- Aizer, A., & Doyle Jr, J. J. (2015). Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges. *The Quarterly Journal of Economics*, 130(2), 759-803.
- Akos, P., & Galassi, J. P. (2004). Middle and high school transitions as viewed by students, parents, and teachers. *Professional School Counseling*, 212-221.
- Allensworth, E. M., & Easton, J. Q. (2005). The On-Track Indicator as a Predictor of High School Graduation. Chicago: Chicago Consortium on School Research. Retrieved from <http://ccsr.uchicago.edu/publications/track-indicator-predictor-high-school-graduation>
- Allensworth, E. M., & Easton, J. Q. (2007). What Matters for Staying On-Track and Graduating in Chicago Public High Schools: A Close Look at Course Grades, Failures, and Attendance in the Freshman Year. Research Report. Chicago: Consortium on Chicago School Research.
- Almås, I., Cappelen, A. W., Sørensen, E. Ø., & Tungodden, B. (2010). Fairness and the development of inequality acceptance. *Science*, 328(5982), 1176-1178.
- Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American statistical Association*, 103(484), 1481-1495.
- Baker, J. A., Derrer, R. D., Davis, S. M., Dinklage-Travis, H. E., Linder, D. S., & Nicholson, M. D. (2001). The flip side of the coin: Understanding the school's contribution to dropout and completion. *School Psychology Quarterly*, 16(4), 406-426.
- Baker, M.L. (2000). *Evaluation of the Truancy Reduction Demonstration Program: Interim Report*. Denver, CO: Colorado Foundation for Families and Children
- Baker, M.L., Sigmon, J.N., & Nugent, M.E. (2001, September). Truancy reduction: Keeping students in school. *Juvenile Justice Bulletin*. Washington DC: Office of Juvenile Justice and Delinquency Prevention.
- Balfanz, R., Herzog, L., & Mac Iver, D.J. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grade schools: Early identification and effective interventions. *Educational Psychologist*, 42, 223-235.
- Barber, B. K., & Olsen, J. A. (2004). Assessing the transitions to middle and high school. *Journal of Adolescent Research*, 19(1), 3-30.
- Beaver, K. M., Wright, J. P., & DeLisi, M. (2011). The racist teacher revisited: Race and social skills in a nationally representative sample of American children. In Frank Columbus (Ed.), *Classrooms: Management, effectiveness, and challenges*, pp. 115-132. New York, NY: Nova Science Publishers
- Benner, A. D. (2011). The transition to high school: Current knowledge, future directions. *Educational Psychology Review*, 23(3), 299.
- Blake J, Butler BR, Smith D. (2015) Challenging Middle-Class Notions of Femininity: The Cause of Black Females' Disproportionate Suspension Rates. In: Losen DJ, ed. Closing the

- School Discipline Gap: Equitable Remedies for Excessive Exclusion. New York, NY: Teachers College Press
- Bostrom, M. (2000). Teenhood: Understanding attitudes toward those transitioning from childhood to adulthood. In S. Bales (Ed.), *Reframing Youth Issues (Working Papers)*. Washington, DC: Frame Works Institute and Center for Communications and Community, UCLA.
- Bowditch, C. (1993). Getting rid of troublemakers: High school disciplinary procedures and the production of dropouts. *Social Problems*, 40(4), 493-509.
- Bowen, N. K., & Bowen, G. L. (1999). Effects of crime and violence in neighborhoods and schools on the school behavior and performance of adolescents. *Journal of Adolescent Research*, 14(3), 319-342.
- Bronfenbrenner, U. (1986). Ecology of the Family as a Context for Human Development: Research Perspectives. *Developmental Psychology*, 22(6), 723-742.
- Bronfenbrenner, U. and Morris, P.A. (1998) The ecology of developmental processes. In W. Damon & R.M. Lerner (Eds), *Handbook of Child Psychology: Theoretical models of human development* (pp. 993-1028). Hoboken, NJ: John Wiley & Sons
- Browne, J. (2003). Derailed: The schoolhouse to jailhouse track. Washington, DC: Advancement Project.
- Bryk, A. S., Sebring, P. B., Allensworth, E., Easton, J. Q., & Luppescu, S. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago, IL: University of Chicago Press.
- Carrell, S. E., & Hoekstra, M. L. (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone's kids. *American Economic Journal: Applied Economics*, 2(1), 211-28.
- Carrell, S. E., Hoekstra, M., & Kuka, E. (2016). The long-run effects of disruptive peers (Working Paper No. 22042). Retrieved from National Bureau of Economic Research website: <http://www.nber.org/papers/w22042>.
- Casey, B. J., Giedd, J. N., & Thomas, K. M. (2000). Structural and functional brain development and its relation to cognitive development. *Biological Psychology*, 54(1), 241-257.
- Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, 29(6), 2156-2160.
- Chicago Public Schools. (2014). *CPS releases outreach plan to strengthen its Suspension and Expulsion Plan, which has already reduced out-of-school suspensions by 36% over three years for high school students*. [Press release]. Retrieved from http://cps.edu/News/Press_releases/Pages/PR1_02_26_2014.aspx
- Chicago Public Schools. (2015). *CPS announces mid-year data: 60 percent fewer out of school suspensions so far this year, 69 percent reduction in expulsions*. [Press release]. Retrieved from http://cps.edu/News/Press_releases/Pages/PR1_3_19_2015.aspx
- Cornell, D., & Huang, F. (2016). Authoritative school climate and high school student risk behavior: A cross-sectional multi-level analysis of student self-reports. *Journal of youth and adolescence*, 45(11), 2246-2259.
- Cornell, D., Shukla, K., & Konold, T. R. (2016). Authoritative School Climate and Student Academic Engagement, Grades, and Aspirations in Middle and High Schools. *AERA Open*, 2(2), 1-18.

- Crenshaw, K. W., Ocen, P., & Nanda, J. (2015). *Black girls matter: Pushed out, overpoliced and underprotected*. New York, NY: African American Policy Forum, Center for Intersectionality and Social Policy Studies.
- Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, 30(3), 418-446.
- Dunbar Jr, C., & Villarruel, F. A. (2002). Urban school leaders and the implementation of zero-tolerance policies: An examination of its implications. *Peabody Journal of Education*, 77(1), 82-104.
- Eaton, S. (2010) Beyond zero tolerance: Creating more inclusive schools by improving neighborhood conditions, attacking racial bias, and reducing inequality. In C. Edley & J. Ruiz de Velasco (Eds). *Changing Places: How Communities will Improve the Health of Boys of Color*. Berkeley, CA: Chief Justice Earl Warren Institute on Race, Ethnicity and Diversity.
- Eberhardt, J. L., & Fiske, S. T. (1994). Affirmative action in theory and practice: Issues of power, ambiguity, and gender versus race. *Basic and Applied Social Psychology*, 15(1-2), 201-220.
- Eden, M. (2017). *School discipline reform and disorder: Evidence from New York City Public Schools, 2012-2016*. New York City: Manhattan Institute
- Eisenberg, N., Carlo, G., Murphy, B., & Van Court, P. (1995). Prosocial development in late adolescence: a longitudinal study. *Child development*, 66(4), 1179-1197.
- Eisenberg, N., Cumberland, A., Guthrie, I. K., Murphy, B. C., & Shepard, S. A. (2005). Age changes in prosocial responding and moral reasoning in adolescence and early adulthood. *Journal of research on adolescence*, 15(3), 235-260.
- Evans, G. (1988). Those loud black girls. *Learning to lose: Sexism and education*, 2, 183-190.
- Fabelo, T., Thompson, M. D., Plotkin, M., Carmichael, D., Marchbanks, M. P., & Booth, E. A. (2011). *Breaking Schools' Rules*. New York, NY: Council of State Governments Justice Center.
- Felner, R. D., Primavera, J., & Cauce, A. M. (1981). The impact of school transitions: A focus for preventive efforts. *American Journal of Community Psychology*, 9(4), 449-459.
- Ferguson, A. A. (2001). *Bad Boys: Public Schools in the Making of Black Masculinity*. Ann Arbor: University of Michigan Press.
- Fiske, S. T. (2000). Stereotyping, prejudice, and discrimination at the seam between the centuries: Evolution, culture, mind, and brain. *European Journal of Social Psychology*, 30(3), 299-322.
- Fordham, S. (1993). "Those Loud Black Girls": (Black) Women, Silence, and Gender" Passing" in the Academy. *Anthropology & Education Quarterly*, 3-32.
- Gilliam, F. D, & Bales, S. N. (2001). Strategic Frame Analysis: Reframing America's Youth. *UCLA: Center for Communications and Community*. Retrieved from <https://escholarship.org/uc/item/5sk7r6gk>
- Gilliam, W.S., Maupin, A.N., Reyes, C.R., Accavitti, M., & Shic, F. (2016). Do early educators' implicit biases regarding sex and race relate to behavior expectations and recommendations of preschool expulsions and suspensions? New Haven: Yale Child Study Center.
- Giroux, H. A., & Purpel, D. E. (1983). *The hidden curriculum and education: Deception or discovery?* Berkeley, CA: McCutchan Pub Corp.

- Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of personality and social psychology*, 70(3), 491-512.
- Goldstein, D. (2017, March 9). Obama Education Rules are Swept Aside by Congress. *The New York Times*, p. A22. Retrieved from <https://www.nytimes.com/2017/03/09/us/every-student-succeeds-act-essa-congress.html>
- Gonzales, R., Richards, K. & Seeley, K. (2002). *Youth Out of School: Linking Absence to Delinquency*. Denver, Colorado: Colorado Foundation for Families and Children.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review*, 102(1), 4-27.
- Gregory, A., Cornell, D., & Fan, X. (2012). Teacher safety and authoritative school climate in high schools. *American Journal of Education*, 118(4), 401-425.
- Gregory, A., Cornell, D., Fan, X., Sheras, P., Shih, T. H., & Huang, F. (2010). Authoritative school discipline: High school practices associated with lower bullying and victimization. *Journal of Educational Psychology*, 102(2), 483-496.
- Gummerum, M., Keller, M., Takezawa, M., & Mata, J. (2008). To give or not to give: Children's and adolescents' sharing and moral negotiations in economic decision situations. *Child Development*, 79(3), 562-576.
- Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior research methods*, 48(1), 400-407.
- Heller, S. B., Shah, A. K., Guryan, J., Ludwig, J., Mullainathan, S., & Pollack, H. A. (2017). Thinking, fast and slow? Some field experiments to reduce crime and dropout in Chicago. *The Quarterly Journal of Economics*, 132(1), 1-54.
- Hemphill, S. A., Heerde, J. A., Herrenkohl, T. I., Toumbourou, J. W., & Catalano, R. F. (2012). The impact of school suspension on student tobacco use: a longitudinal study in Victoria, Australia, and Washington State, United States. *Health Education & Behavior*, 39(1), 45-56.
- Henry, K.L & Thornberry, T.P. (2010). Truancy and escalation of substance use during adolescence. *Journal of Studies on Alcohol and Drugs*, 71, 115-124.
- Henry, K.L. (2010). Skipping school and using drugs: A brief report. *Drugs: Education, Prevention and Policy*, 17(5), 650-657.
- Henry, K.L., & Huizinga, D.H. (2007b). Truancy's effect on the onset of drug use among urban adolescents placed at risk. *Journal of Adolescent Health*, 40(4), e9-e17.
- Hook, J. G., & Cook, T. D. (1979). Equity theory and the cognitive ability of children. *Psychological Bulletin*, 86(3), 429-445.
- Hopkins, B. (2002). Restorative justice in schools. *Support for Learning*, 17(3), 144-149.
- Huang, F. L. (2016). Do Black students misbehave more? Investigating the differential involvement hypothesis and out-of-school suspensions. *The Journal of Educational Research*, 1-11.
- Huang, F. L., & Cornell, D. G. (2017). Student attitudes and behaviors as explanations for the Black-White suspension gap. *Children and youth services review*, 73, 298-308.
- Huff, C., & Tingley, D. (2015). "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research & Politics*, 2(3), 1-12.

- Jacob, B. and Lefgren, L. (2003). "Are Idle Hands the Devil's Workshop? Incapacitation, Concentration and Juvenile Crime." *American Economic Review*, 93(5): 1560-1577
- Jia, Y., Konold, T. R., & Cornell, D. (2016). Authoritative school climate and high school dropout rates. *School Psychology Quarterly*, 31(2), 289-303.
- Kim, C. Y., Losen, D. J., & Hewitt, D. T. (2010). *The school-to-prison pipeline: Structuring legal reform*. NYU Press.
- Kinsler, J. (2011). Understanding the black-white school discipline gap. *Economics of Education Review*, 30(6), 1370-1383.
- Kirk, D. S., & Papachristos, A. V. (2011). Cultural mechanisms and the persistence of neighborhood violence. *American Journal of Sociology*, 116(4), 1190-1233.
- Kirk, D. S., & Sampson, R. J. (2013). Juvenile arrest and collateral educational damage in the transition to adulthood. *Sociology of Education*, 86(1), 36-62.
- Klein, R. (2014, July 31). Teachers May not Want to Reduce Number of Out-of-school Suspensions, Survey Suggests. *Huffington Post*. Retrieved from https://www.huffingtonpost.com/2014/07/31/school-superintendents-discipline-survey_n_5636077.html
- Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75(1), 83-119.
- Kohlberg, L. (1983). *Essays in moral development, vol. II: The psychology of moral development*. San Francisco: Harper & Row.
- Koonce, J. B. (2012). "Oh, Those Loud Black Girls!": A Phenomenological Study of Black Girls Talking with an Attitude. *Journal of Language and Literacy Education*, 8(2), 26-46.
- Kupchik, A. (2010). *Homeroom security: School discipline in an age of fear*. New York: New York University Press.
- Lacoe, J., & Steinberg, M. (In Press). Rolling back zero tolerance: The effect of discipline policy reform on suspension usage and student outcomes. *Peabody Journal of Education*.
- Levin, H. S., Culhane, K. A., Hartmann, J., Evankovich, K., Mattson, A. J., Harward, H., ... & Fletcher, J. M. (1991). Developmental changes in performance on tests of purported frontal lobe functioning. *Developmental Neuropsychology*, 7(3), 377-395.
- Liberman, A. M., Kirk, D. S., & Kim, K. (2014). Labeling effects of first juvenile arrests: Secondary deviance and secondary sanctioning. *Criminology*, 52(3), 345-370.
- Losen, D. J., & Gillespie, J. (2012). *Opportunities suspended: The disparate impact of disciplinary exclusion from school*. Los Angeles, CA: The Center for Civil Rights Remedies.
- Losen, D. J., & Martinez, T. E. (2013). *Out of school & off track: The overuse of suspensions in American middle and high schools*. Los Angeles, CA: The UCLA Center for Civil Rights Remedies at The Civil Rights Project.
- Losen, D. L. & Skiba, R.J. (2010). *Suspended education: Urban middle schools in crisis*. The Civil Rights Project at UCLA and the Southern Poverty Law Center.
- Luppescu, S. and Ehrlich, S. (2012). *The New and Improved CCSR Rasch Measurement Model Primer*. Chicago: University of Chicago Consortium on School Research. Retrieved from http://consortium.uchicago.edu/sites/default/files/uploads/survey/Overview%20of%20Analyses_0.pdf

- MacDonald, H. (Summer 2012). "Undisciplined: The Obama Administration Undermines Classroom Order in Pursuit of Phantom Racism." *City Journal*.
- MacGillivray, H., Medal, M., & Drake, C. (2008). Zero tolerance policies: A precarious balance between school safety and educational opportunity for all. In K.G. Welner & W.C. Chi (Eds), *Current Issues in Education Policy and the Law*, (pp. 191-212). Charlotte, NC: Information Age Publishing, Inc.
- Martin, K. A. (1998). Becoming a gendered body: Practices of preschools. *American Sociological Review*, 494-511.
- Mendez, L. M. R., Knoff, H. M., & Ferron, J. M. (2002). School demographic variables and out-of-school suspension rates: A quantitative and qualitative analysis of a large, ethnically diverse school district. *Psychology in the Schools*, 39(3), 259-277.
- Mendis, T. (2017, May 26). Dress code violation gains national attention: Principal receives threats. WCNC. Retrieved from <http://www.wcnc.com/news/local/dress-code-violation-gains-national-attention-principal-receives-threats/443247003>
- Merten, D. E. (1997). The meaning of meanness: Popularity, competition, and conflict among junior high school girls. *Sociology of Education*, 175-191.
- Monahan, K. C., VanDerhei, S., Bechtold, J., & Cauffman, E. (2014). From the school yard to the squad car: school discipline, truancy, and arrest. *Journal of Youth and Adolescence*, 43(7), 1110-1122.
- Moore, M.H., Petrie, C.V., Braga, A.A. & McLaughlin, B.L. (Eds) (2003). *Deadly lessons: Understanding lethal school violence*. Washington, DC: National Academy Press
- Morris, E. W. (2005). "Tuck in that shirt!" Race, class, gender, and discipline in an urban school. *Sociological Perspectives*, 48(1), 25-48.
- Morris, E. W. (2007). "Ladies" or "loudies"? Perceptions and experiences of black girls in classrooms. *Youth & Society*, 38(4), 490-515.
- Morris, M. (2018). *Pushout: The Criminalization of Black Girls in Schools*. New York: The New Press.
- Morrison, G. M., & D'Incau, B. (1997). The web of zero-tolerance: Characteristics of students who are recommended for expulsion from school. *Education and Treatment of Children*, 20(3), 316-335.
- Morrison, G. M., Anthony, S., Storino, M. H., Cheng, J. J., Furlong, M. J., & Morrison, R. L. (2001). School expulsion as a process and an event: Before and after effects on children at risk for school discipline. *New Directions for Student Leadership*, 2001(92), 45-71.
- Mowen, T., & Brent, J. (2016). School discipline as a turning point: The cumulative effect of suspension on arrest. *Journal of research in crime and delinquency*, 53(5), 628-653.
- Newman, B. M., Newman, P. R., Griffen, S., O'Connor, K., & Spas, J. (2007). The relationship of social support to depressive symptoms during the transition to high school. *Adolescence*, 42(167), 441.
- Nolan, K. (2011). *Police in the hallways: Discipline in an urban high school*. Minneapolis: University of Minnesota Press.
- Okonofua, J. A., & Eberhardt, J. L. (2015). Two strikes: Race and the disciplining of young students. *Psychological science*, 26(5), 617-624.
- Owens, E. (2015, May 15). Oakland Kids who Cuss Teachers out no Longer Need to Fret about Getting Suspended. *The Daily Caller*. Retrieved from

- <http://dailycaller.com/2015/05/15/oakland-kids-who-cuss-teachers-out-no-longer-need-to-fret-about-getting-suspended/>
- Perez, J. (2015, February 25). Teachers Complain about Revised CPS Discipline Policy. *Chicago Tribune*. Retrieved from <http://www.chicagotribune.com/news/ct-cps-discipline-concerns-met-20150225-story.html>
- Perry, B. L., & Morris, E. W. (2014). Suspending progress: Collateral consequences of exclusionary punishment in public schools. *American Sociological Review*, 79(6), 1067-1087.
- Phillips, M. (2011). Parenting, time use, and disparities in academic outcomes. In G.J. Duncan & R.J. Murnane (Eds). *Whither Opportunity*, pp 207-228. New York: Russell Sage Foundation
- Purdie-Vaughns, V., & Eibach, R. P. (2008). Intersectional invisibility: The distinctive advantages and disadvantages of multiple subordinate-group identities. *Sex Roles*, 59(5-6), 377-391.
- Raby, R. (2005). Polite, well-dressed and on time: Secondary school conduct codes and the production of docile citizens. *Canadian Review of Sociology*, 42(1), 71-91.
- Rios, V. M. (2011). *Punished: Policing the lives of Black and Latino boys*. New York: NYU Press.
- Rocque, M. (2010). Office discipline and student behavior: Does race matter?. *American Journal of Education*, 116(4), 557-581.
- Rocque, M., & Paternoster, R. (2011). Understanding the antecedents of the "school-to-jail" link: The relationship between race and school discipline. *The Journal of Criminal Law and Criminology*, 633-665.
- Roderick, M. (2003). What's happening to the boys? Early high school experiences and school outcomes among African American male adolescents in Chicago. *Urban Education*, 38, 538-607.
- Rosenblat, J. (2015, March 17). Schools Hope Changes in Policy will Bridge 'Discipline Gap.' *US News and World Report*. Retrieved from <https://www.usnews.com/news/articles/2015/03/17/schools-hope-changes-in-policy-will-bridge-the-discipline-gap>
- Sameroff, A. J. (1983). Developmental systems: Contexts and evolution. In P.H. Mussen & W. Kessen (Eds.), *Handbook of child psychology: Vol. 1. History, theory and methods* (pp. 237-294). New York, NY: Wiley.
- Sameroff, A. J. (1994). Developmental systems and family functioning. In R. D. Parke & S.G. Kellam (Eds.), *Exploring family relationships with other social contexts* (pp 199-214). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sameroff, A. J. & Fiese, B.H. (1990). Transactional regulation and early intervention. In S.J. Meisels & J.P. Shonkoff (Eds.), *Handbook of early childhood intervention* (pp 119-149). New York, NY: Cambridge University Press.
- Sampson, R. J., & Laub, J. H. (1997). A life-course theory of cumulative disadvantage and the stability of delinquency. *Developmental theories of crime and delinquency*, 7, 133-161.
- Schwerdt, G., & West, M. R. (2013). The impact of alternative grade configurations on student outcomes through middle and high school. *Journal of Public Economics*, 97, 308-326.

- Seidman, E., Lawrence Aber, J., Allen, L., & French, S. E. (1996). The impact of the transition to high school on the self-system and perceived social context of poor urban youth. *American Journal of Community Psychology*, 24(4), 489-515.
- Simmons, R. (2002). *Odd girl out: The hidden culture of aggression in girls*. New York: Houghton Mifflin Harcourt.
- Skiba, R. J., & Rausch, M. K. (2006). Zero tolerance, suspension, and expulsion: Questions of equity and effectiveness. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 1063–1089). Mahwah, NJ: Erlbaum
- 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.
- 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.
- Skiba, R., & Peterson, R. (1999). The dark side of zero tolerance: Can punishment lead to safe schools? *Phi Delta Kappan*, 80(5), 372-382.
- Skiba, R., Simmons, A., Staudinger, L. Rausch, M. Dow, G., & R. Feggins. (2003). Consistent Removal: Contributions of School Discipline to the School-Prison Pipeline. Paper presented to the School to Prison Pipeline Conference, Harvard University. 16-17 May. Bloomington, IN: Indiana Education Policy Center at Indiana University
- Steinberg, L. (2005). Cognitive and affective development in adolescence. *Trends in Cognitive Sciences*, 9(2), 69-74.
- Steinberg, L. (2007). Risk taking in adolescence: New perspectives from brain and behavioral science. *Current Directions in Psychological Science*, 16(2), 55-59.
- Steinberg, M. P., & Laco, J. (2017). What Do We Know About School Discipline Reform? *Education Next*, 17(1), p.43-52.
- Stevens, W.D., Sartain, L., Allensworth, E.M., and R. Levenstein. (2015). Discipline Practices in Chicago Schools: Trends in the Use of Suspensions and Arrests. Chicago: The University of Chicago Consortium on Chicago School Research.
- Sutter, M. (2007). Outcomes versus intentions: On the nature of fair behavior and its development with age. *Journal of Economic Psychology*, 28(1), 69-78.
- Sweeten, G. (2006). Who will graduate? Disruption of high school education by arrest and court involvement. *Justice Quarterly*, 23(4), 462-480.
- 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.
- Turiel, E. (2005). Resistance and subversion in everyday life. In L. Nucci (Ed.), *Conflict, contradiction, and contrarian elements in moral development and education* (pp. 3-20). Mahwah, NJ: Erlbaum
- U.S. Department of Education Office of Civil Rights (2016). "2013-2014 Civil Rights Data Collection: A First Look". Retrieved from <http://www2.ed.gov/about/offices/list/ocr/docs/2013-14-first-look.pdf>

- US Department of Education and US Department of Justice. (2011). *Attorney General Holder, Secretary Duncan announce effort to respond to school-to-prison pipeline by supporting good discipline practices* [Press release]. Retrieved from <https://www.justice.gov/opa/pr/attorney-general-holder-secretary-duncan-announce-effort-respond-school-prison-pipeline>.
- US Department of Education and US Department of Justice. (2014). "Dear colleague" Letter on the Nondiscriminatory Administration of School Discipline. Retrieved from: <https://www2.ed.gov/about/offices/list/ocr/letters/colleague-201401-title-vi.html#ftn13>
- US Department of Education Office for Civil Rights (2016, June). *2013-2014 Civil Rights Data Collection: A First Look*. Washington, DC: US Department of Education. Retrieved from <https://www2.ed.gov/about/offices/list/ocr/docs/2013-14-first-look.pdf>
- US Department of Education Office for Civil Rights (2018, April). *2015-2016 Civil Rights Data Collection: School Climate and Safety*. Washington, DC: US Department of Education. Retrieved from <https://www2.ed.gov/about/offices/list/ocr/docs/school-climate-and-safety.pdf>
- Vavrus, F., & Cole, K. (2002). "I Didn't Do Nothin'": The Discursive Construction of School Suspension. *The Urban Review*, 34(2), 87-111.
- Waldron, L. M. (2011). "Girls Are Worse" Drama Queens, Ghetto Girls, Tomboys, and the Meaning of Girl Fights. *Youth & Society*, 43(4), 1298-1334.
- Watanabe, T. (2013, May 14). LA Unified Bans Suspension for "Willful Defiance." *Los Angeles Times*. Retrieved from <http://articles.latimes.com/2013/may/14/local/la-me-laUSD-suspension-20130515>
- Watanabe, T. and Blume, H. (2015, November 7). Why some LAUSD Teachers are Balking at a New Approach to Discipline Problems. *Los Angeles Times*. Retrieved from <http://www.latimes.com/local/education/la-me-school-discipline-20151108-story.html>
- Wehlage, G., & Rutter, R. (1986). Dropping out: How much do schools contribute to the problem? *The Teachers College Record*, 87(3), 374-392.
- Welch, K., & Payne, A. A. (2010). Racial threat and punitive school discipline. *Social Problems*, 57(1), 25-48.
- Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology*, 7(2), 131-149.
- Welsh, W. N. (2000). The effects of school climate on school disorder. *The Annals of the American Academy of Political and Social Science*, 567(1), 88-107.
- Wiley, S. A., Slocum, L. A., & Esbensen, F.A. (2013). The unintended consequences of being stopped or arrested: An exploration of the labeling mechanisms through which police contact leads to subsequent delinquency. *Criminology*, 51(4), 927-966.
- Willert, T. (2015a, September 2). Oklahoma City District Sees Lower Suspension Rate, Official Says. *The Oklahoman*. Retrieved from <http://newsok.com/article/5443960>
- Willert, T. (2015b, September 23). Oklahoma City District Suspends Fewer Students, Report Shows. *The Oklahoman*. Retrieved from <http://newsok.com/article/5448743>
- Wright, J. P., Morgan, M. A., Coyne, M. A., Beaver, K. M., & Barnes, J. C. (2014). Prior problem behavior accounts for the racial gap in school suspensions. *Journal of Criminal Justice*, 42(3), 257-266.

- Wu, S.C., Pink, W., Crain, R. & O, Moles. (1982). Student Suspension: A Critical Reappraisal. *The Urban Review*. 14(4), 245-303.
- Wun, C. (2016). Against captivity: Black girls and school discipline policies in the afterlife of slavery. *Educational policy*, 30(1), 171-196.
- Wyn, J. (2005). Youth in the Media: Adult stereotypes of young people. In A. Williams and C. Thurlow (Eds), *Talking Adolescence: Perspectives on Communication in the Teenage Years*. New York, NY: Peter Lang
- Zeedyk, M. S., Gallacher, J., Henderson, M., Hope, G., Husband, B., & Lindsay, K. (2003). Negotiating the transition from primary to secondary school: Perceptions of pupils, parents and teachers. *School Psychology International*, 24(1), 67-79.

Appendix Tables

Appendix Table 2.1. Treatment Check

Variable	(1)	(2)	(3)	(4)	(5)
	White Male	Black Male	White Female	Black Female	F-Stat p-value
Likelihood that the student is...					
black	-0.6422 (0.6614)	0.5246 (0.9483)	-0.2021 (0.9196)	0.3151 (1.0149)	0.000
from a low-income neighborhood	-0.0873 (0.9548)	0.1569 (1.1268)	0.1091 (0.9864)	-0.1835 (0.9012)	0.173
Observations	64	65	62	62	253

Appendix Table 3.1. Suspension Prediction

Dependent Variable: Ever Received an Out-of-School Suspension in Freshmen Year (Logit Coefficients)

VARIABLES	(1) OSS
male	0.019 (0.025)
Other language at home	-0.486*** (0.043)
black	0.773*** (0.058)
Hispanic	0.366*** (0.058)
IEP	-0.098** (0.035)
Emotionally Disabled	0.611*** (0.098)
Free/reduced lunch	0.442*** (0.052)
8 th grade absence rate	0.016*** (0.002)
7 th grade absence rate	0.011*** (0.002)
8 th grade GPA	-0.459*** (0.023)
7 th grade GPA	-0.248*** (0.022)
8 th grade reading test score	-0.070** (0.023)
8 th grade math test score	-0.013 (0.026)
7 th grade reading test score	-0.098*** (0.023)
7 th grade math test score	0.009 (0.026)
Any 8 th grade arrest charges	0.400*** (0.031)
Constant	-1.038*** (0.088)
Observations	59,810
Schools	NO

Appendix Table 3.2. School-Level Parallel Trends Tests

VARIABLES	(1) # OSS/Stud (m)	(2) OSS Length (m)	(3) Ever OSS (a)	(4) # ISS/Stud (m)	(5) ISS Length (m)	(6) Ever ISS (a)
High Suspending * SY11	0.026 (0.201)	0.837 (0.551)	0.026 (0.042)	-0.720 (0.572)	-0.421 (0.772)	0.069 (1.315)
High Suspending * SY12	0.125 (0.311)	0.677 (0.750)	0.047 (0.056)	-0.983 (0.659)	-0.969 (0.809)	-0.546 (0.679)
High Suspending * SY14	-0.379* (0.147)	-1.230* (0.464)	-0.071* (0.031)	-0.160 (0.329)	0.012 (0.421)	-0.169 (0.524)
High Suspending * SY15	-0.449 (0.228)	-2.036** (0.635)	-0.103* (0.045)	0.441 (0.897)	0.514 (1.000)	-0.074 (1.113)
Observations	145	145	145	145	145	145
R-squared	0.409	0.530	0.542	0.372	0.354	0.213
Number of schlid	29	29	29	29	29	29
Covs	x	x	x	x	x	x
Year Fixed Effects	x	x	x	x	x	x

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, *

p<0.05

**Appendix 3.2. School-Level Parallel Trends Tests
(continued)**

	(7)	(8)	(9)	(10)	(11)
VARIABLES	Attendance	Academic	Classroom Order	School Climate	Total Arrest Charges Per Stud
High Suspending * SY11	0.077 (0.074)	-0.095 (0.081)	0.169 (0.114)	-0.022 (0.096)	-0.152* (0.062)
High Suspending * SY12	-0.052 (0.058)	-0.110 (0.088)	0.053 (0.158)	-0.024 (0.092)	-0.066 (0.062)
High Suspending * SY14	0.166** (0.059)	-0.018 (0.079)	0.087 (0.098)	0.106 (0.078)	-0.059 (0.037)
High Suspending * SY15	0.116 (0.067)	-0.112 (0.077)	0.480** (0.135)	0.110 (0.085)	-0.110* (0.044)
Observations	145	145	145	145	145
R-squared	0.385	0.406	0.344	0.188	0.371
Number of schlid	29	29	29	29	29
Covs	x	x	x	x	x
Year Fixed Effects	x	x	x	x	x

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Appendix Table 3.3. Average Suspension Length from Misconduct vs. Attendance file

Data Source	OSS Length		ISS Length	
	Misconduct	Attendance	Misconduct	Attendance
Pre-Policy High Susp School Mean	4.356	4.252	1.814	2.609
High Susp School* SY14	-1.629*** (0.410)	-1.581*** (0.378)	0.392 (0.367)	-0.031 (0.385)
High Susp School* SY15	-2.522*** (0.514)	-2.368*** (0.481)	0.945 (0.829)	0.064 (1.074)
Observations	145	145	145	145
R-squared	0.518	0.534	0.343	0.210
Number of schlid	29	29	29	29
Covs	All	All	All	All

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Appendix Table 3.4. Missing Survey Responses

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Share missing MVMS	Share male missing MVMS	Share IEP missing MVMS	Share ED missing MVMS	Share other language @ home missing MVMS	Share black missing MVMS	Share Hispanic missing MVMS	Share FRL missing MVMS
Pre-Policy High Susp School Mean	0.445	0.468	0.533	0.599	0.142	0.449	0.329	0.439
High Susp School* SY14	-0.025 (0.096)	-0.020 (0.094)	0.101 (0.112)	0.240 (0.190)	-0.018 (0.128)	0.032 (0.107)	-0.071 (0.150)	-0.024 (0.094)
High Susp School* SY15	0.027 (0.076)	0.040 (0.082)	0.011 (0.089)	-0.219 (0.166)	0.319 (0.160)	0.061 (0.084)	0.084 (0.175)	0.032 (0.074)
Observations	145	145	145	145	145	145	145	145
R-squared	0.130	0.157	0.165	0.262	0.181	0.103	0.080	0.130
Covs	x	x	x	x	x	x	x	x
Year Fixed Effects	x	x	x	x	x	x	x	x
Number of schlid	29	29	29	29	29	29	29	29

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Appendix Table 3.5. Parallel Trends Test for High Risk Students Only

VARIABLES	(1) # OSS (m)	(2) OSS Length (m)	(3) Ever OSS (a)	(4) # ISS (m)	(5) ISS Length (m)	(6) Ever ISS (a)
High Suspending * SY11	0.010 (0.328)	0.911 (0.940)	0.047 (0.065)	-0.193 (0.417)	0.327 (0.643)	-0.035 (0.136)
High Suspending * SY12	0.000 (0.292)	0.735 (0.861)	0.045 (0.049)	-0.050 (0.326)	0.363 (0.565)	0.035 (0.125)
High Suspending * SY14	-0.232 (0.218)	-0.589 (0.709)	0.050 (0.038)	0.541 (0.383)	1.028 (0.534)	0.005 (0.056)
High Suspending * SY15	-0.847* (0.393)	-2.323* (0.947)	-0.113 (0.071)	1.243 (0.971)	1.723 (1.151)	-0.002 (0.116)
Observations	6,202	6,202	6,202	6,202	6,202	6,202
R-squared	0.103	0.112	0.075	0.074	0.057	0.022
Number of schlid	29	29	29	29	29	29
Model	High Studs Only	High Studs Only	High Studs Only	High Studs Only	High Studs Only	High Studs Only

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Appendix Table 3.5. Parallel Trends Test for High Risk Students Only (continued)

VARIABLES	(7) Attendance Index	(8) Academic Index	(9) Classroom Order	(10) School Climate Index	(11) Total Arrest Charges
High Suspending * SY11	0.023 (0.128)	-0.201 (0.131)	0.117 (0.102)	-0.018 (0.069)	-0.008 (0.092)
High Suspending * SY12	-0.010 (0.079)	-0.154 (0.096)	0.136 (0.117)	0.020 (0.060)	0.006 (0.064)
High Suspending * SY14	0.017 (0.097)	-0.115 (0.111)	-0.147 (0.106)	0.126 (0.087)	-0.034 (0.086)
High Suspending * SY15	0.055 (0.074)	-0.196 (0.103)	0.217 (0.173)	0.052 (0.074)	-0.112 (0.073)
Observations	6,202	6,202	3,307	3,307	6,202
R-squared	0.301	0.230	0.024	0.028	0.239
Number of schlid	29	29	29	29	29
Model	High Studs Only	High Studs Only	High Studs Only	High Studs Only	High Studs Only

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Appendix Table 3.6. Parallel Trends Test for Lower Risk Students Only

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	# OSS (m)	OSS Length (m)	Ever OSS (a)	# ISS (m)	ISS Length (m)	Ever ISS (a)
High Suspending * SY11	0.101 (0.246)	0.370 (0.475)	0.022 (0.075)	-0.192 (0.197)	-0.140 (0.226)	-0.072 (0.071)
High Suspending * SY12	-0.087 (0.135)	0.030 (0.222)	-0.008 (0.044)	-0.074 (0.141)	0.015 (0.179)	-0.036 (0.051)
High Suspending * SY14	-0.224** (0.070)	-0.448 (0.234)	-0.074** (0.026)	0.100 (0.138)	0.165 (0.149)	-0.023 (0.060)
High Suspending * SY15	-0.406** (0.136)	-0.951* (0.449)	-0.164** (0.045)	0.090 (0.239)	0.123 (0.256)	-0.040 (0.051)
Observations	19,736	19,736	19,736	19,736	19,736	19,736
R-squared	0.064	0.048	0.060	0.041	0.041	0.036
Number of schlid	29	29	29	29	29	29
Model	High Studs Only	High Studs Only	High Studs Only	High Studs Only	High Studs Only	High Studs Only

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Appendix Table 3.6. Parallel Trends Test for Lower Risk Students Only (continued)

VARIABLES	(7) Attendance Index	(8) Academic Index	(9) Classroom Order	(10) School Climate Index	(11) Total Arrest Charges
High Suspending * SY11	-0.017 (0.051)	-0.116 (0.122)	0.093 (0.086)	0.001 (0.096)	-0.021 (0.020)
High Suspending * SY12	-0.049 (0.062)	-0.081 (0.066)	0.172** (0.058)	-0.017 (0.045)	-0.032 (0.017)
High Suspending * SY14	0.061 (0.068)	0.057 (0.053)	-0.027 (0.160)	-0.012 (0.085)	-0.045* (0.021)
High Suspending * SY15	0.072 (0.038)	0.068 (0.093)	0.238 (0.144)	0.044 (0.064)	-0.078* (0.028)
Observations	19,736	19,736	15,474	15,474	19,736
R-squared	0.314	0.302	0.024	0.018	0.048
Number of schlid	29	29	29	29	29
Model	High Studs Only	High Studs Only	High Studs Only	High Studs Only	High Studs Only

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Appendices

Appendix A. Name Selection

The names were generated with the help of David Figlio at Northwestern University. He provided us with the list after analyzing over 2 million birth certificates from large state in the United States. This data set provides the name, race and maternal education for each birth certificate. He provided us with a list of names that appeared in the data set with moderate frequency, are at least 90% white or 90% black, are clearly gender-identifying, and have identical maternal education levels. We selected three names for testing for each of the following four groups: black boy, black girl, white boy and white girl. These produced the following sets:

- a. Relatively lower educated – mean maternal education 11.49 to 11.62 years:
 1. Black boys: Keyshawn, Deonte
 2. Black girls: Keondra, Latasha
 3. White boy: Coty, Codey, Tylor
 4. White girl: Misty, Katlynn
- b. Relatively higher educated – mean maternal education 13.05 to 13.15 years:
 1. Black boy: Tahj
 2. Black girl: Imani
 3. White boy: Seth
 4. White girl: Lila

Although the data tells us that these names are distinctly white/black, male/female, and matched on maternal education, our survey experiment relies on survey participants' *perceptions* of each name's race, gender, and income status. In other words, it is possible that an objectively distinct black female high income names (based on birth certificates in a certain state) may not be perceived as very black, female, or high income by respondents. We use Mturk to test the validity of these names with experimental vignettes described in sections below. We recruited

about 20 individuals per name. Each study participants was presented with the following scenario and questions:

Imagine yourself as a teacher at a junior high school located in a middle income neighborhood in your region of the country. The average student teacher ratio is 26 students to every one teacher at the school. The stories were collected through a previous study in which we asked actual teachers to describe a typical incident involving a misbehaving student. Please read the story carefully and answer the questions that follow as though you were the actual teacher in the class.

You have a student named _____ in your class. _____ is consistently disrupting the class environment by strolling around the classroom at random intervals, getting tissues from the tissue box multiple times during a 50 minute class, throwing items away constantly; in general, _____ circulates around the room and up and down the rows to see what other students are doing, tries to get other students' attention, and disrupts the flow of the lecture or activity the class was participating in.

Please answer the following 7 questions:

1. How likely is it that _____ is a male?
2. How likely is it that _____ is from a low-income neighborhood?
3. How likely is it that _____ is black?
4. How severe was _____ behavior?
5. To what extent is _____ hindering you from maintaining order in your class?
6. How irritated do you feel by _____?
7. How severely should _____ be disciplined?

Each question was answered on a scale from 1 to 7. We focused primarily on questions one to three. The mean responses for each name are presented in Table 1 below.

Table 1

Name	Mean male	Mean low-income	Mean black
Keyshawn (n=23)	5.13 (1.65)	4.61 (1.41)	4.35 (.168)
Deonte (n=21)	5.57 (1.56)	4.05 (1.56)	4.48 (1.65)
Tahj (n=25)	5.40 (1.52)	4.00 (1.41)	3.92 (1.19)
Jalen (n=16)	6.44 (0.73)	4.38 (1.20)	4.60 (1.18)
Keondra (n=12)	4.25 (1.74)	4.58 (1.32)	5.08 (1.26)

Latasha (n=26)	3.04 (1.65)	4.69 (1.35)	4.54 (1.55)
Imani (n=10)	4.40 (1.69)	3.60 (1.02)	3.80 (1.40)
Symone (n=49)	4.82 (1.81)	4.61 (1.54)	4.41 (1.44)
Codey (n=23)	5.43 (1.47)	4.43 (1.31)	3.91 (1.59)
Tylor (n=22)	5.59 (1.11)	4.36 (1.07)	4.00 (1.31)
Seth (n=22)	5.68 (1.49)	4.05 (1.40)	3.64 (1.26)
Colby (n=25)	5.08 (1.71)	3.64 (1.55)	3.52 (1.50)
Misty (n=24)	2.83 (1.62)	3.60 (1.47)	3.68 (1.71)
Katlynn (n= 23)	3.52 (1.69)	4.00 (1.38)	4.14 (1.21)
Taryn (n=22)	4.61(1.44)	4.50 (1.47)	4.73 (1.16)
Lila (n=25)	3.44 (1.65)	4.48 (1.10)	4.24 (1.14)

Surprisingly, our names did not reveal large differences in perceived race for this first group of names. We decided to supplement our names list with those from a recent kindergarten discrimination paper (Gilliam, Maupin, Reyes, Accavitti and Shic, 2016). This study used Deshawn, Jake, Latoya and Emily. We tested these names on Mturk using the same vignettes as described above. The results are presented below in Table 2.

Table 2.

Name	Mean male	Mean low-income	Mean black
Deshawn (n=55)	5.45 (1.63)	4.47 (1.56)	4.62 (1.43)
Jake (n=50)	5.71 (1.50)	4.26 (1.47)	4.16 (1.27)
Latoya (n=55)	3.95 (2.04)	4.07 (1.36)	4.56 (1.32)

Emily (n=49)	3.61 (2.03)	4.14 (1.48)	3.74 (1.28)
--------------	-------------	-------------	-------------

We looked across all of the tested names to see which names matched the best on gender and socioeconomic status, but also varied the most on perceived race. Thus, the final set of names is:

1. Black boy: Deonte
2. Black girl: LaToya
3. White boy: Codey
4. White girl: Emily

Appendix B. Survey Scenarios

Scenario One

LaToya came in late to your classroom during Test day. You asked for her tardy pass. She didn't respond. You asked again for her to give you her tardy pass. She slammed it on your desk. Then, while the class was taking the test, LaToya made a lot of noise stomping to her desk. And 20 minutes later, she threw her pencil at the trashcan.

Scenario Two

LaToya is upset because you "bother" her and "yell" at her because she "wants to sit quietly and do nothing". And she says that you should just leave her alone. So you give LaToya reading assignments and just busy work. But she doesn't do anything you give her because she says you're "crazy".

Appendix C. Modifications to the Student Code of Conduct

- 1) Network Chief or designee approval is required for suspensions of pre-kindergarten through second grade students
- 2) Students in third through twelfth grade can be assigned out-of-school suspensions only if a student's attendance at school endangers others, or if a student causes chronic or extreme interruption to other students' participation in school activities and prior interventions have been utilized
- 3) Administrators are required to develop a plan to support students if they are suspended for three or more days
- 4) Out-of-school suspensions are removed as an available consequence for responding to Group 2 and first-time Group 3 behaviors
- 5) Number of suspension days permitted for repeated Group 3 and Group 4 and Group 5 behaviors is lowered.
- 6) Administrators have discretion to assign shorter-term suspensions for Groups 5 and 6 behaviors
- 7) Removes the Group 5 "persistent defiance" infraction since it is too vague.
- 8) Allows referral of students to district-level intervention programs in lieu of expulsion for Groups 5 and 6 behaviors
- 9) Emphasizes the use of interventions and consequences that address the cause of the inappropriate behavior

Appendix D. School Climate Survey Questions

The *positive peer environment*, *personal safety*, *school safety concerns* (reverse coded), *student-teacher connection*, and *school-connectedness* indices were created by UChicago CCSR using Rasch analysis (Luppescu & Ehrlich, S. 2012). *Positive peer environment* asks students to report on how much students put each other down, help each other learn, get along together, and treat each other with respect. Negative items are reverse scored so an increase in this index represents an increase in positive peer environments. *Personal safety* measures the extent students feel safe in hallways, bathrooms, outside around school, traveling between home and school, and in their classrooms. Higher scores mean students feel safer. *School safety concern* examines how students view the school culture in terms of safety. It asks students to report on how much they worry about crime and violence at school, how often they think other students are teased/picked on, and how often students are threatened/bullied. This measure has been reverse coded for my analysis so that an increase in the school safety index represents an increase in students' sense of safety. *Student-teacher connection* analyzes students' trust in and respect for their teachers. Higher scores represent feeling a stronger connection with their teachers. Finally, the *school connection* index measures the extent to which students feel like they are a valued member of their school community. Higher scores on this index represent a stronger school connection.

Survey questions wording

Classroom Disorder

How much do you agree with the following statements about your class?

- 1) This class is out of control

Strongly disagree, Disagree, Agree, Strongly agree

Ever Bullied

In the last year...

- 1) Have you ever been afraid someone was going to bully you at school?
- 2) Have you ever been afraid someone was going to bully you online?
- 3) Have you ever been afraid someone might bully you because they thought you were gay, lesbian, bisexual, or transgender?

Strongly disagree, Disagree, Agree, Strongly agree

Positive Peer Environment

How much do you disagree or agree with the following statements about students in your school? Most students in my school...

- 1) Like to put others down
- 2) Help each other learn
- 3) Don't get along together very well
- 4) Treat each other with respect
- 5)

Strongly disagree, Disagree, Agree, Strongly agree

Personal Safety

How safe do you feel...

- 1) In the hallways of the school?
- 2) In the bathrooms of the school?
- 3) Outside around the school?
- 4) Traveling between home and school?
- 5) In your classes?

Not safe, Somewhat safe, Mostly safe, Very safe

School Safety Concern

How much do you agree with the following statements about your school:

- 1) I worry about crime and violence in this school
- 2) Students at this school are often teased or picked on
- 3) Students at this school are often threatened or bullied

Strongly disagree, Disagree, Agree, Strongly agree

Teacher Connection

- 1) When my teachers tell me not to do something, I know he/she has a good reason
- 2) I feel safe and comfortable with my teachers at this school
- 3) My teachers always keep their promises
- 4) My teachers will always listen to students' ideas
- 5) My teachers treat me with respect

Strongly disagree, Disagree, Agree, Strongly agree

School Connection

How much do you agree with the following statements about your school:

- 1) I feel like a real part of my school
- 2) People here notice when I'm good at something
- 3) Other students in my school take my opinions seriously
- 4) People at this school are friendly to me
- 5) I'm included in lots of activities at school

Strongly disagree, Disagree, Agree, Strongly agree

Appendix E. Sample Sizes by School Suspension and Student Risk

	High Suspending School	Low Suspending School
High Risk Students	4178	2024
Lower Risk Students	3463	16273

Appendix F. Student Interview Protocol

Background

A. School

- How long have you been at this school?
- Tell me about this school
- Tell me about the kids here
- Tell me about the teachers here
- How did you choose this school?

Suspension Process

A. Personal experiences

- What, if anything, did you know about the suspension process before the first time you were suspended?
- Do you think school suspensions are useful? Why?
- Describe the first time you were ever suspended.
 - Tell me what happened
 - What was the suspension process like? (Take me from the incident, through suspension, and back to school)
 - How did the process make you feel?
 - Do you think it was fair to give you a suspension? Why or why not?
- [If this is not the first suspension, ask:] Describe the most recent time you were suspended
 - Tell me what happened
 - What was the suspension process like? (Take me from the incident, through suspension, and back to school)
 - How did the process make you feel?
 - What changes, if any, would you make to the suspension process? Or school discipline policies more generally?

B. Family Experiences

- What did your mom/guardian say about the suspension?
- What did they have to do because you got suspended?
- Who was around the house when you were suspended?
 - Is that different or same as usual?

Returning to school

- What's it like returning to school?
 - What were you most looking forward to?
 - What were you most worried about?
- How did the suspension process make you feel?
- Why do you think schools use suspensions?
- Do you think it achieved its purpose/was successful in your case?

Concluding thoughts

- Can you think of someone that you feel should have been suspended but was not? Tell me about this instance.
- If you were the principal of a school, would you keep suspensions as a consequence? Why or why not?
- Based on what we've already discussed about the suspension policy and school discipline policies, is there anything you feel I should know that I haven't asked you about?
- Are there any questions you think I should have asked that I haven't?
- work week?)