

Rolling Back Zero Tolerance: The Effect of Discipline Policy Reform on Suspension Usage and Student Outcomes

Johanna Lacoë
Researcher, Mathematica Policy Research
505 14th Street, Suite 800, Oakland, CA 94612

Matthew P. Steinberg
Assistant Professor, University of Pennsylvania
3700 Walnut Street, Philadelphia, PA 19104

October 31, 2017

Keywords: discipline policy, out-of-school suspensions, quasi-experimental methods, zero tolerance reform

JEL classification: I21, I28

The authors thank Peter Schochet, John MacDonald and Matthew Chingos for feedback on earlier versions of the paper and conference participants at the Association for Public Policy Analysis and Management (APPAM) and Association for Education Finance and Policy (AEFP) and seminar participants at the University of Arkansas, Department of Education Reform Lecture Series for helpful comments and discussions. We thank Cameron Anglum for excellent research assistance and Jennifer Moore for editorial assistance. We thank the Pennsylvania Department of Education for providing access to achievement and demographic data, and Tonya Wolford and Rachel Holzman from the School District of Philadelphia for helpful discussions. Finally, we thank the Fordham Institute for ongoing support of this project. Authors contributed equally. Lacoë (corresponding author) may be contacted at JLacoë@mathematica-mpr.com. Steinberg may be contacted at steima@upenn.edu.

Abstract

Beginning in the early 1990s, states and districts enacted zero-tolerance discipline policies that relied heavily on out-of-school suspensions. Recently, districts nationwide have revised these policies in favor of more tempered disciplinary responses. In 2012–2013, Philadelphia reformed its discipline policy to limit suspensions for non-violent student misconduct and granted principals greater discretion in responding to more serious occurrences of student misconduct. Employing a difference-in-differences approach, we find that Philadelphia’s reform resulted in a modest decline in suspensions for non-violent infractions in the year of reform; however, total suspensions remained unchanged while serious incidents of student misconduct increased. Further, the truancy rate increased and district math and ELA achievement declined following the policy reform. These findings should inform policymakers and practitioners on the implications of district-level reforms for suspension usage and the potential consequences for student outcomes.

Introduction

Beginning in the early 1990s, many states and school districts, particularly those serving children from low-income families, enacted stricter zero-tolerance discipline policies that relied heavily on the use of out-of-school suspensions (Curran 2016). Critiques of zero-tolerance policies, which employ major consequences such as suspension for even minor infractions, have surfaced in the academic literature and the popular press (Shah 2013; Skiba 2008). Articles in the *New York Times* have challenged policies designed to increase police presence in schools and the efficacy of using suspensions and arrests to improve school safety (Eckholm 2013; Yee 2013). Several national student advocacy groups joined the Dignity in Schools Campaign calling for a moratorium on suspensions (Shah 2012), and the American Academy of Pediatrics' Committee on School Health has been critical of the overuse of school suspensions (American Academy of Pediatrics 2003). Much of the criticism of strict zero-tolerance policies centers on the use of out-of-school suspension for lower-level, nonviolent student behavior. This is in light of evidence that nearly half of all suspensions issued in California schools during the 2011-2012 school year were for 'willful defiance,' a category of student misconduct that includes a student's refusal to remove a hat, turn off a cellphone or failure to wear the school's uniform (Watanabe 2013).

In response, states and districts across the country have begun to revise zero-tolerance policies in favor of more tempered disciplinary responses that aim to keep students in the classroom. As of May 2015, 22 states and the District of Columbia had revised their laws to require or encourage schools to limit the use of exclusionary discipline practices, implement supportive (i.e., non-punitive) discipline strategies that rely on behavioral interventions such as restorative justice and school-wide positive behavioral interventions and supports (SWPBIS), and provide support services such as counseling, dropout prevention, and guidance services for at-risk students (i.e., students with academic and/or mental health disabilities). By the 2015–2016 school year, 23 of the nation's 100 largest school districts implemented policy reforms that limit the use of suspensions or require less punitive discipline strategies (Steinberg & Lacoë 2017). A recent survey by the American Association of School Administrators found that over half of the 464 districts surveyed had revised their student code of conduct to include changes in the use of nonpunitive responses to misbehavior, out-of-school suspensions and expulsions, and the length of suspensions (American Association of School Administrators 2013). Changes to discipline codes of conduct in many large, urban school districts reflect a growing shift away from strict zero-tolerance policies (Baker 2012; Mezzacappa 2012; Paslay 2012; Taube 2013; Watanabe 2013). For example, public school systems in Baltimore, Chicago, and Cincinnati have changed their discipline policies to decrease the use of suspensions and incorporate school-wide positive behavioral programs (New York City School-Justice Partnership Task Force 2013). In Los Angeles, Baltimore, and Rochester, NY, school districts are relying more heavily on data systems to track schools with high suspension and high crime rates. Most of these cities have documented decreases in suspensions following these interventions and policy changes (New York City School-Justice Partnership Task Force 2013).

Despite increasing efforts nationwide to limit the use of exclusionary discipline practices, there remains limited empirical evidence on the impact of recent district-level policy reforms (Steinberg & Lacoë 2017). Understanding the districtwide impact of such reforms is critical for a number of reasons. First, local public actors (such as school principals) typically have relative autonomy from organizational authorities (such as district policymakers) to implement policy

reforms (Lipsky 2010). Misalignment between district administrators' goals for the reform and the ways in which school leaders implement the reform at the local level may limit the efficacy of district-level policy efforts to reduce the use of suspensions. Research has highlighted the importance of school principals' perspectives on discipline in determining suspension rates (Skiba et al. 2014). Second, if district-level policy reforms are not coupled with additional resources to support local implementation, schools may be unable to meet district policymakers' intended goals. Indeed, few district-level discipline reforms provide clear guidance or sufficient resources to school leaders to support alternative methods to prevent and respond to student misbehavior. As a result, the efficacy of district-level policy reforms to achieve their intended goals remains an open empirical question. Further, very little is known about the potential unintended consequences of such policy reforms on other important outcomes, including school absences, student misconduct and academic achievement.

In the 2012–2013 school year, the School District of Philadelphia (SDP) made dramatic changes to its discipline code of conduct. The SDP's discipline policy reform emphasized intervention rather than suspension for less serious student misconduct and gave school principals greater discretion around their response to more serious disciplinary infractions. In this paper, we leverage longitudinal district-level data to answer the questions: (1) Did Philadelphia's discipline policy reform reduce suspension usage? (2) Did the discipline policy reform affect school absences, student misconduct and academic achievement?

Given that Philadelphia's decision to reform its discipline policy was not made at random, we employ a quasi-experimental panel difference-in-differences approach to estimate the impact of the policy reform on suspension usage, student misconduct, truancy, and academic achievement. We find consistent evidence that Philadelphia's code of conduct reform did not reduce the overall use of suspensions in the post-reform years. While we do observe a short-term decrease in non-violent conduct infractions which were the target of the reform, these declines do not persist over the post-policy period. We also find that more serious incidents of student misconduct increased in the wake of the reform, and that unexcused absences (truancy) increased and math and ELA achievement declined districtwide in the wake of the policy reform. Even in the absence of the random assignment of district-level policy reform, the use of district-level panel data allows us to generate a more rigorous estimate of the impact of discipline policy reform than currently exists in the nascent literature.

In the proceeding sections, we begin by summarizing the existing literature on school climate, safety, and discipline. Next, we detail the policy change that occurred in the School District of Philadelphia. With this background in place, we describe our data, sample, and empirical approach, and then present our results. The final section provides a discussion of the policy implications of the findings.

Related Literature

School climate and safety shape the educational lives of students. Feeling unsafe at school negatively affects student achievement (Bryk, Sebring, Allensworth, Luppescu, and Easton 2010; Lacoé 2016), and students who are victims of harassment attend school less frequently and have lower levels of school engagement, impeding their school participation and cognitive and social development (Bowen and Bowen, 1999; MacMillan and Hagan 2004; Payne, Gottfredson, and

Gottfredson 2003). The role that school climate and safety play in student achievement depends on a number of classroom-level mechanisms. For example, children who are physically and verbally abusive in the classroom divert teachers' attention away from teaching (Bowen and Bowen, 1999), and more experienced, higher-performing teachers are more likely to leave schools with substantial disciplinary problems (Allensworth, Ponisciak, and Mazzeo 2009; Payne et al. 2003; Smith and Smith 2006; Kraft, Marinell and Yee 2016). Safety is of particular concern in urban schools located in neighborhoods with concentrated poverty and high crime rates (Bowen, Bowen, and Ware 2002; Felson, Liska, South, and McNulty, 1994; Steinberg et al. 2011; Welsh, Greene, and Jenkins, 1999). In the long term, attendance at unsafe schools disrupts a child's educational and occupational attainment, affecting later life outcomes such as economic status, labor force participation, and earnings (MacMillan and Hagan 2004; Schreck and Miller 2003).

One strategy for maintaining safe and orderly schools is to remove disruptive students through out-of-school suspension. Yet, teachers and students report feeling less safe in schools with higher suspension rates compared with schools serving students from similar neighborhoods and with similar academic and socioeconomic characteristics but with lower suspension rates (Bradshaw, Sawyer, and O'Brennan 2009; Lacoë 2015; Steinberg et al. 2011). Schools with higher suspension rates have greater teacher attrition and turnover (Mendez, Knoff, and Ferron 2002) and are associated with lower levels of student achievement and higher dropout rates (Rausch and Skiba 2006; Rich 2012). At the student level, school disciplinary actions decrease student academic achievement (Lacoë and Steinberg 2017) and are correlated with an increase in the probability of dropping out of school (Peguero and Bracy 2015). A recent study, which leverages student-level panel data and controls for time-invariant student characteristics, finds that suspensions decrease math and ELA test scores and the probability that suspended students meet state proficiency standards; further, exposure to suspensions has small negative effects on suspended students' non-offending peers (Lacoë and Steinberg 2017). Exclusionary discipline policies are also associated with greater levels of student fear at school and loss of instructional time, and there is little evidence that such policies reduce school violence (American Psychological Association Zero Tolerance Task Force 2008; Hoffman 2014; Schreck and Miller 2003).

Suspensions are also disproportionately centered on economically disadvantaged students, students with emotional and behavioral disorders, and minority students (Osher, Bear, Sprague, and Doyle 2010). Nationwide, African American students were more than four times as likely to be suspended during the 2009–2010 school year as their White counterparts, and nearly twice as many students with disabilities were suspended than students without disabilities (Losen and Gillespie 2012). In Texas between 2000 and 2002, African American students were 31 percent more likely to receive a disciplinary action than White and Hispanic students, and approximately three quarters of all special education students were either suspended or expelled (Fabelo et al. 2011). Further, school structure, principal perspectives on discipline, and students' perceptions of disciplinary fairness have implications for racial differences in achievement (Arum 2003; Kupchik and Ellis 2008; Gregory, Cornell and Fan 2011; Skiba, Chung, Baker, Sheya, and Hughes 2014). Research highlights large gaps in performance between Black and White students in high schools where students view the disciplinary scheme to be unfair and lenient, in contrast to no racial differences in performance in schools where students perceive discipline to be strict and fair (Arum 2003).

While recent policy reforms aim to address these concerns by reducing the use of out-of-school suspension, particularly for less serious (i.e., non-violent) conduct infractions, there is little empirical evidence documenting the impact of rolling back exclusionary discipline policies on suspension usage and school climate and performance. The only evidence comes from district-level policy reform in Chicago Public Schools, which changed its district code of student conduct in 2012-2013 to reduce the length of out-of-school suspensions for more serious infractions from ten to five consecutive days. Evidence from Chicago indicates that reductions in suspension length increased learning time for offending students (i.e., those students subject to longer suspensions for more serious misconduct) without adversely affecting the achievement of their non-offending peers or the overall school climate (Mader et al. 2016).

In this paper, we provide new evidence on the effect of a policy reform that targeted suspensions for low-level conduct infractions. Efforts to reduce suspensions for non-violent infractions may be a promising avenue for district-level discipline reforms, given new research that suggests that suspending students for low-level misconduct imposes significant costs on their academic performance (Lacoe and Steinberg 2017). At the same time, since low-level student misconduct is both more prevalent and therefore more likely to shape the classroom context – by creating disorder and disrupting instruction – than less frequent and more serious infractions, district level reform efforts may have unintended consequences on student behavior, engagement, and achievement. Up to this point, however, the consequences of limiting suspensions for conduct infractions remain unknown.

Policy Context

Zero-tolerance discipline policies, an outgrowth of state and federal drug enforcement policies, emerged in the late 1980s and early 1990s in response to an increasing frequency of violent student behaviors (Hanson 2005). Amid growing national concern with school violence, the federal government under the Clinton Administration passed the Gun-Free Schools Act, authorized under the Improving America's Schools Act of 1994 (later repealed and then re-enacted under the No Child Left Behind Act of 2001). Under this law, states and local school districts enacted zero-tolerance policies of varying degrees, with some policies punishing students in possession of weapons, look-alike weapons, or objects construed as weapons (Hanson 2005). In an extreme example, in 2001, a 10-year-old Florida student found a small knife in her lunchbox that her mother placed there for cutting an apple. The student immediately handed over the knife to her teacher; however, she was expelled from school for possessing a weapon (American Psychological Association Zero Tolerance Task Force 2008).

According to Hanson (2005), state and local zero-tolerance policies have failed to make three important distinctions. First, the policies treated all children equally, making no distinction in the child's age or grade and, therefore, imposing the same punishment—suspension or expulsion—irrespective of whether the child was a kindergarten student or a high school senior. Second, students could be expelled for what was construed as a weapon—such as a water pistol—or what appeared to be an illicit substance—such as Tylenol. Third, no consideration was made regarding the educational achievement, psychological development, or learning needs of a student, so that all students—regardless of individual circumstances—received the same punishment. Ultimately, zero-tolerance policies replaced the authority and discretion of teachers and school leaders to interpret and address individual disciplinary infractions.

In August 2012, the School District of Philadelphia held a summit of school principals to discuss preventative strategies to improve school safety. In addition, a private Philadelphia foundation funded a 2-year fellowship to develop a “school safety and climate strategy” for the district, based on collaboration among the SDP; the Philadelphia Departments of Behavioral Health, Police, and Human Services; and student and parent representatives (Snyder 2012). These reform efforts culminated in changes to the SDP’s Code of Student Conduct.

Beginning in September 2012, the SDP adopted a revised code of conduct that emphasized intervention rather than suspension or disciplinary transfers (i.e., transferring students across schools for disciplinary reasons); gave principals greater discretion in handling more serious disciplinary cases; and prevented students from receiving out-of-school suspensions for less severe, conduct infractions (Mezzacappa 2012). Under the revised code of conduct, students were no longer to be removed from school for failing to follow classroom rules or using profane or obscene language or gestures; instead, the maximum allowable punishment changed from out-of-school suspension (for 1–3 days) to in-school intervention (see Table A1 for a summary of the changes to the SDP’s Code of Student Conduct). In the final pre-policy year (i.e. 2011-2012), 15 percent of all suspensions among grade 3–12 students were for failure to follow classroom rules and/or disruption and 10 percent were for the use of profane/obscene language or gestures. Among students in grades 3–12 receiving at least one out-of-school suspension in 2011-2012, 22 percent were suspended for failure to follow classroom rules and/or disruption and 15 percent were suspended for the use of profane/obscene language or gestures. For other conduct infractions such as public displays of affection, inappropriate use of electronic devices and forgery of an adult’s signature, the policy change required in-school intervention as a first response, with out-of-school suspension to be used only as a last resort.

For more serious offenses—such as theft, harassment and bullying, breaking and entering, robbery, or extortion—maximum punishments were changed from expulsion to suspension (potentially paired with assignment to a disciplinary school, or another type of transfer). The code reforms also gave school administrators greater discretion in handling student misbehavior, as well as determining the consequences for more serious student misconduct. In doing so, Philadelphia’s policy reform not only aimed to reduce the use of suspension for more minor student misconduct, but also provided school leaders with the type of decision-making authority typically absent under strict, zero-tolerance approaches to school discipline.

Data and Sample

To explore the consequences of Philadelphia’s 2012 discipline policy reform, we constructed a district-level panel dataset for the 2005-2006 through 2014-2015 school years. We merged district-level demographic data from the U.S. Department of Education’s National Center for Education Statistics (i.e., Common Core of Data [CCD]) with district-level enrollment, achievement, and suspension data from the Pennsylvania Department of Education (PDE).

District-level demographic data include the proportion of a district’s students receiving free or reduced-price lunch; the proportion of students identified as English language learners; the proportion of students identified as disabled and receiving individualized education plans; and the proportion of students who are African American or Hispanic (i.e., minority students). We supplement district-level demographic data from CCD with district enrollment (all students

enrolled in traditional, non-charter, public schools) from PDE. Charter school students are excluded because charter schools have their own discipline policies and are not subject to the district's student code of conduct.

Our discipline data include counts of serious incidents and counts of out-of-school suspensions (OSS). Beginning in the 2005-2006 school year, PDE has required all Pennsylvania school districts to report the total annual count of OSS and the total annual count of serious incidents.¹ District-level serious incident data is reported in aggregate and includes all occurrences of serious student misconduct and does not include lower-level, conduct infractions. The Pennsylvania Department of Education defines a serious incident as a “specific act or offense involving one or more victims or offenders... These include, but are not limited to, any behavior that violates a school’s educational mission or climate of respect, or jeopardizes the intent of the school to be free of aggression against persons or property, drugs, weapons, disruptions, and disorder. Examples are incidents involving acts of violence, possession of a weapon, or the possession, use or sale of a controlled substance, alcohol, or tobacco by any person on school property, at school-sponsored events, and on school transportation to and from school.” (Pennsylvania Information Management System 2015).

All suspension data are reported on an incident basis, so that a student who receives an OSS multiple times in the same academic year will be counted multiple times in the total OSS count. PDE also requires districts to disaggregate OSS data into six mutually exclusive categories describing the suspension reason; these reasons include academic, conduct, drugs and alcohol, tobacco, violence, or weapons. For example, an academic OSS is for academic reasons (e.g., cheating), while a conduct OSS is for violations of the district’s student code of conduct. Importantly, PDE requires districts to report more serious offenses (such as drug and alcohol possession, violent behavior and weapons possession) separately from conduct incidents (see Table A2 for PDE incident reporting categories). Given that Philadelphia’s 2012 reform focused on eliminating OSS for lower-level, non-violent infractions—such as failure to follow classroom rules or using profane language or gestures—while discouraging OSS for other conduct offenses (such as public displays of affection, inappropriate use of electronic devices and forgery of adult’s signature), this level of disaggregation allows insight into changes in OSS that may be centered on lower-level, nonviolent violations of the student code of conduct apart from more serious offenses.

Outcome Measures

To examine whether the discipline policy reform affected suspension use, we examine three OSS outcomes. The first is *aggregate (or total) OSS*, which is inclusive of all OSS across the six PDE reporting categories (academic, conduct, drugs and alcohol, tobacco, violence, and weapons). Since we are principally concerned with whether Philadelphia’s policy reform affected suspension usage for lower-level offenses, we construct a second OSS measure – *conduct OSS* – which includes just conduct OSS and excludes OSS for all other reasons. If Philadelphia’s policy reform had an effect on suspension usage, we expect the effect to be concentrated on conduct OSS. To examine whether the effect of the policy reform was

¹ Suspension and incident data retrieved from <https://www.safeschools.state.pa.us>

concentrated on conduct OSS, we construct a third OSS measure – *nonconduct OSS* – which includes OSS for all reasons except for lower-level conduct and academic offenses. All OSS outcomes are calculated on a per-capita basis.²

While changes to the student code of conduct may be designed to reduce suspensions, the extent to which discipline policy reform influences school climate more broadly remains an open question. Policy-induced reductions in OSS may lead to more disruption and worse school climate if students who formerly would have been suspended from school now remain in the classroom. Alternatively, if students and teachers perceive the new code of conduct to be fairer than the previous policy, and teachers and school leaders are given greater discretion to respond to more serious disciplinary infractions, policy reforms may improve student behavior (Arum 2003; Mitchell and Bradshaw 2013). We therefore extend our analysis to consider the extent to which the policy reform affected other aspects of school climate and student outcomes, including student behavior, absences, and achievement.

The policy reform gave principals greater autonomy to more actively address serious incidents of student misbehavior, both through intervention and de-escalation as well as the prevention of more serious misconduct. To do so, principals were encouraged to utilize alternative approaches to school discipline such as restorative justice, approaches which address school culture, the sources of student misconduct, and disciplinary responses to that misconduct. Research shows that some of these approaches, including positive behavioral interventions and supports (PBIS) (Bradshaw, Mitchell, and Leaf 2010) and restorative justice (Anyon et al. 2016), are effective at reducing disciplinary referrals and improving student behavior. However, in Philadelphia, school leaders were given little guidance about which alternative strategies to use and few additional resources to implement new discipline approaches.³ Therefore, whether the policy reform affected occurrences of more serious student misconduct is an empirical question. To estimate the effect of the policy reform on student behavior, we create a measure of *serious incidents*, which is the per-capita count of all serious incidents reported by each school district in a given school year.

The direction of the relationship between the policy reform and truancy (i.e., absences) is also, *ex ante*, unclear. On the one hand, if school climate improves following the policy change, truancy rates may improve. For example, evidence suggests that schoolwide efforts to limit the use of exclusionary discipline practices and implement supportive (i.e., non-punitive) discipline strategies that rely on behavioral interventions (e.g., school-wide PBIS) can improve school attendance (Freeman, Simonsen, McCoach, Sugai, Lombardi, and Horner 2016). On the other hand, if reducing OSS leads to lower levels of school safety, some students may stay home more often. Notably, Lacoë (2016) finds that middle school students in New York City who feel unsafe at school are more likely to be absent from school. We measure each district's *truancy rate*, which PDE defines as the share of a school's students who are truant, or who meet a

² Per capita out-of-school suspensions (OSS) is calculated by dividing the total number of out-of-school suspensions in district *i* during school year *t* by total enrollment (K–12) in district *i* during school year *t*.

³ In conversations with Rachel Holzman, Philadelphia's Deputy Chief of Student Rights and Responsibilities, the authors learned that few additional resources (such as in-school counselors or support staff) were provided to school principals to support the policy change (authors' correspondence on March 23, 2015).

specified level of truancy within a specified period of time – three days of absence from compulsory education without a lawful excuse.⁴ Investigating the consequence of discipline policy reform on truancy rates is important, since unexcused absences from school are associated with lower educational attainment and delinquent behavior in youth (Baker, Sigmon and Nugent 2001).

Finally, the policy change may affect academic achievement through two channels: first, by improving achievement among students who would have otherwise been suspended (and instead remain in the classroom and receive more instruction); and second, by reducing achievement among these students' peers if classrooms are more disorderly following the policy change. While suspensions are negatively related to achievement (Raffaele Mendez 2003; Lacoë and Steinberg 2017), evidence suggests that the presence of disruptive and disorderly students leads to adverse achievement consequences for their classroom peers. Specifically, students who are regularly tardy disrupt the classroom learning environment and negatively affect the achievement of their peers (Gottfried 2014). Further, evidence elsewhere finds that students who were exposed to Hurricane Katrina evacuees with more discipline problems themselves experienced increased school absences and discipline problems, suggesting a short-term adverse effect on peer behavior (Imberman, Kugler and Sacerdote 2012). In the longer-term, exposure to disruptive peers during elementary school reduces high school achievement and college enrollment (Carrell, Hoekstra and Kuka 2016). In addition, discipline policy reform of the kind in Philadelphia required teachers to incorporate disruptive students back into the learning community. If teachers are not equipped with classroom management and instructional strategies to manage the inclusion of these potentially more disruptive students, instruction time for the classroom as a whole may suffer, negatively affecting achievement. We construct a district-level *math (ELA) achievement* measure equal to the average share of a district's students who were proficient or advanced on the math (ELA) portion of the Pennsylvania System of School Assessment, weighted by the number of tested students in a subject*grade*year cell.⁵

Sample

We include 483 (of the 500) Pennsylvania school districts in our analytic sample, including Philadelphia and 482 districts from which we construct comparison groups. To limit the sample to districts with plausible OSS, we drop 15 districts in which the difference between the minimum and maximum per-capita OSS over the study period (i.e., 2005–2006 through 2014–2015) exceeded 2 standard deviations of total per-capita OSS (a difference of 0.5 per-capita OSS, or 50 suspensions per 100 students). See Figure A1 for the distribution of the maximum difference in OSS for all districts. In addition, two districts that merged (Center Area School District and Monaca School District merged to form the Central Valley School District in the 2009–2010 school year) and one district that did not operate any public schools (Bryn Athyn School District) during the study period were also excluded. We also drop any Pennsylvania

⁴ Source: Pennsylvania Department of Education, http://www.education.pa.gov/Documents/K-12/Safe%20Schools/School%20Attendance%20Improvement%20Truancy%20Reduction_Rev_February%202015.pdf

⁵ PSSA data were retrieved from <http://www.education.pa.gov/Data-and-Statistics/Pages/default.aspx#.ViVThX6rSUK>

school districts that revised their discipline code of conduct (to reduce zero-tolerance disciplinary policies) during the study period.⁶

Table 1 summarizes district demographic characteristics and OSS for Philadelphia and the 482 comparison districts in both the pre-reform and post-reform periods. In the pre-reform period, Philadelphia, by far the largest district in the state, enrolled approximately 167,000 students, compared with an average of approximately 3,000 students in all other districts. In addition, Philadelphia serves a predominantly minority and economically disadvantaged student population, with approximately three quarters of its students receiving free/reduced-price lunch. Philadelphia's students were comparatively lower achieving; approximately half of Philadelphia's students were academically proficient in math and reading in the pre-reform period, while approximately 75 percent of students in all other districts met or exceeded state academic standards in math and reading.

In addition to demographic differences between Philadelphia and the rest of Pennsylvania's school districts, Philadelphia also had much higher levels of serious incidents and OSS over the pre-reform period than the average Pennsylvania district (see Table 1). In the pre-reform period, the serious incident rate was more than twice as high in Philadelphia (seven serious incidents per 100 students) compared with all other Pennsylvania districts (three serious incidents per 100 students). Total OSS was five times greater in Philadelphia (30 OSS per 100 students) than in all other districts (six OSS per 100 students), with most of total OSS made up of conduct (i.e., lower-level) offenses (25 conduct OSS per 100 students in Philadelphia and four conduct OSS per 100 students in all other districts).

⁶ We did find that a few Pennsylvania school districts changed their code of conduct policy in the post-reform period. For example, Pittsburgh Public Schools revised its student code of conduct beginning in the 2014–2015 school year to focus on progressive discipline practices designed to improve student behavior and keep students in school by reducing punishments for low-level offenses that previously received OSS (source: <http://csgjusticecenter.org/youth/media-clips/pittsburgh-school-board-approves-revised-code-of-conduct/>). Due to large variation in OSS rates during the study period, Pittsburgh is one of the 15 districts removed from the analytical sample, therefore policy changes in Pittsburgh do not affect the current estimates. Woodland Hills School District (located near Pittsburgh) revised its student code of conduct in 2015–2016 to reduce zero-tolerance policies that the district viewed as disproportionately affected minority students (source: <http://www.post-gazette.com/news/education/2015/07/31/Woodland-Hills-recognized-for-discipline-policies/stories/201507310136>). The 2015–2016 school year falls outside of the post-reform period considered in this paper.

Table 1. District Characteristics

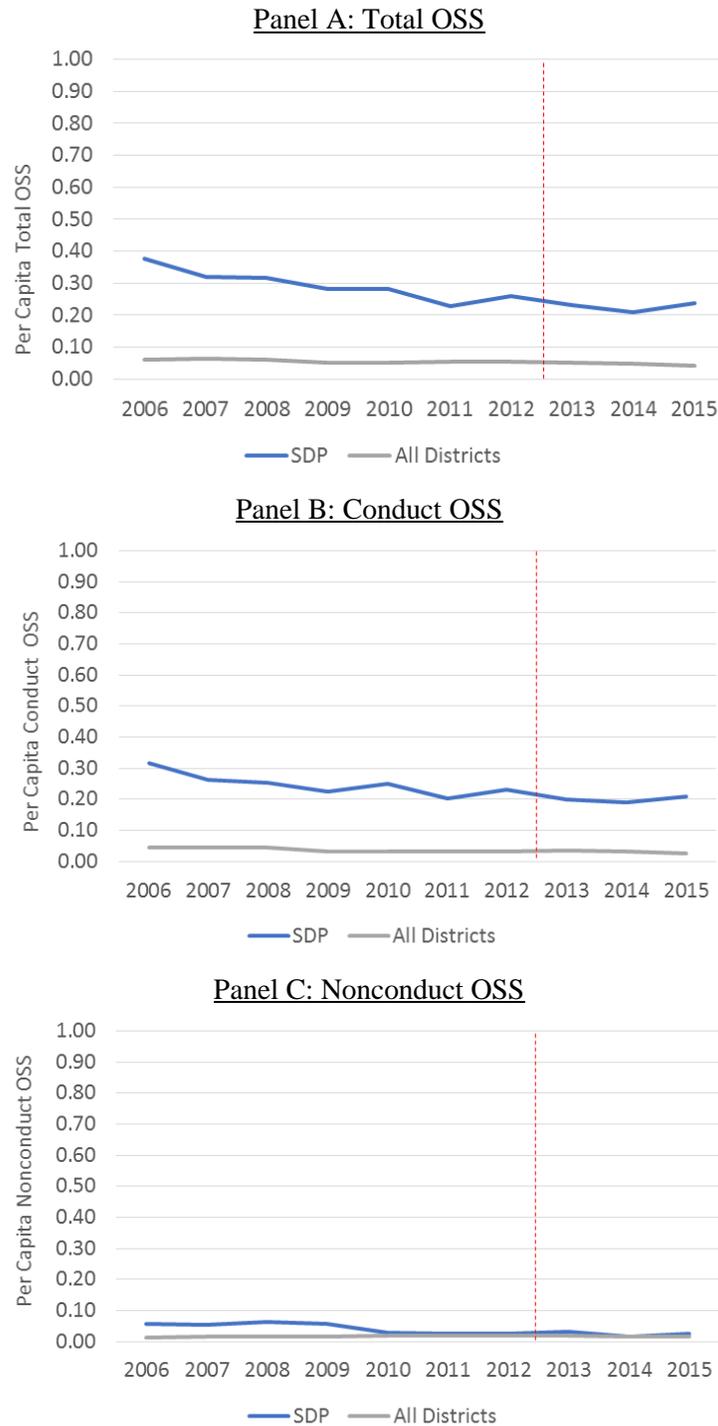
District characteristic	Pre-reform period (2005/2006–2011/2012)		Post-reform period (2012/2013–2014/2015)	
	Philadelphia	All Districts (n=482)	Philadelphia	All Districts (n=482)
Enrollment	166,709.6 (7,409.1)	2,999.1 (2,597.9)	138,604.3 (4,895.3)	2,885.8 (2,553.7)
Free and reduced-price lunch	0.74 (0.030)	0.32 (0.160)	0.83 (0.024)	0.39 (0.165)
English language learners	0.07 (0.003)	0.01 (0.020)	0.09 (0.009)	0.01 (0.021)
Disabled	0.15 (0.010)	0.17 (0.035)	0.16 (0.020)	0.17 (0.036)
Minority	0.78 (0.040)	0.09 (0.138)	0.72 (0.022)	0.10 (0.144)
Math achievement	50.4 (5.73)	76.3 (9.06)	38.0 (15.81)	66.6 (18.10)
Reading achievement	45.4 (5.07)	73.8 (9.21)	41.8 (5.79)	70.9 (11.13)
Serious incidents	0.066 (0.015)	0.028 (0.029)	0.041 (0.006)	0.024 (0.024)
Truancy rate	0.343 (0.055)	0.034 (0.058)	0.377 (0.062)	0.048 (0.083)
Total OSS	0.295 (0.047)	0.056 (0.068)	0.226 (0.015)	0.048 (0.063)
Conduct OSS	0.249 (0.036)	0.038 (0.059)	0.199 (0.009)	0.031 (0.054)
Nonconduct OSS	0.045 (0.017)	0.018 (0.017)	0.025 (0.008)	0.017 (0.016)

Note. Mean (standard deviation) reported. Math (reading) achievement represents the percentage of tested students meeting state proficiency standards on the Pennsylvania System of School Assessment. OSS and serious incidents reported on a per-capita (per-pupil) basis. All other district characteristics reported in proportions. *Total OSS* includes all out-of-school suspensions (academic, conduct, drugs and alcohol, tobacco, violence, and weapons). *Conduct OSS* includes suspensions for only conduct reasons. *Nonconduct OSS* includes out-of-school suspensions for all reasons except for lower-level conduct and academic offenses. There are 482 comparison districts.

Trends in Out-of-School Suspensions

Figure 1 shows OSS in Philadelphia and average OSS across all Pennsylvania comparison districts during the study period. The figure reveals some important trends. First, total OSS in Philadelphia was falling steadily in the pre-reform period (i.e., 2005-2006 through 2011-2012), while mean total OSS among all other Pennsylvania districts remained fairly stable (see Panel A). Specifically, during the pre-reform period, total per-capita OSS fell by 37 percent in Philadelphia – from 0.38, or 38 suspensions per 100 students in 2005-2006 to 0.26, or 26 suspensions per 100 students in 2011-2012; among all other Pennsylvania districts, total per-capita OSS declined by 11 percent – from 0.06 in 2005-2006 to 0.05 in 2011-2012. For conduct OSS, the target of Philadelphia’s policy reform, a similar trend emerges. Namely, conduct OSS in Philadelphia declined from 0.32 to 0.23 during the pre-reform period, while average conduct OSS among all other Pennsylvania districts declined over this same period from 0.045 to 0.034 (see Panel B).

Figure 1. OSS Trends, Philadelphia and All Other Pennsylvania Districts



Note. The year 2006 refers to the 2005–2006 school year. *SDP* refers to the School District of Philadelphia; *All Districts* include all 482 comparison districts. The vertical line indicates the first year (2012–2013) of Philadelphia’s policy reform. *Total OSS* is the per-capita number of out-of-school suspensions, inclusive of all suspension types (academic, conduct, drugs and alcohol, tobacco, violence, and weapons); *Conduct OSS* is the per-capita number of conduct out-of-school suspensions; and *Nonconduct OSS* is the per-capita number of out-of-school suspensions for reasons related to drugs and alcohol, tobacco, violence, and weapons (but exclusive of suspensions due to violations of the student code of conduct or academic reasons).

We also observe significant post-reform differences in OSS. Relative to the pre-reform period, total OSS in Philadelphia declined by 27 percent in the post-reform period – from 0.295, on average, in the pre-reform period to 0.226, on average, in the post-reform period. For all other Pennsylvania districts, average total OSS declined by 15 percent – from 0.056 to 0.048. Further, conduct OSS declined in Philadelphia by 5 conduct suspensions per 100 students – from 0.249, on average, in the pre-reform period to 0.199, on average, in the post-reform period; for all other Pennsylvania districts, average conduct OSS declined by less than one per 100 students – from 0.038 in the pre-reform period to 0.031 in the post-reform period (see Table 1). While these post-reform declines in OSS are not evidence that Philadelphia’s code of conduct reform decreased OSS, they do suggest that there were potentially meaningful changes in the use of exclusionary discipline practices in the wake of the reform. We next explore the extent to which Philadelphia’s policy reform led to changes in suspension usage and other district outcomes.

Empirical Approach

Given that OSS is declining in Philadelphia during the pre-reform period (see Figure 1), our aim is to distinguish any post-reform decline in OSS that is due to Philadelphia’s discipline policy reform from any post-reform change in OSS that may be due to district-specific secular time trends. To do so, we implement a quasi-experimental panel difference-in-differences approach which accounts for the secular decline in OSS rates in Philadelphia. We discuss this approach below.

As seen in Figure 1, the parallel trends assumption that underlies a simple difference-in-differences approach is clearly violated, given the downward sloping OSS trend in Philadelphia and the relatively constant pre-reform trend in all other Pennsylvania districts. However, we employ a panel difference-in-differences (DD) strategy that enables us to control for district-specific linear time trends. We specify the model as:

$$(1) \text{OSS}_{it} = \beta_0 + \beta_1 T_{it} + \mathbf{X}_{it} \boldsymbol{\Gamma} + \theta_i + \lambda_t + \gamma_i t + \varepsilon_{it}$$

where OSS_{it} is a measure of suspensions (per-capita) for district i in school year t ; T_{it} is an indicator variable that takes on a value of one for Philadelphia in the post-reform period (i.e., 2012–2013 through 2014–2015 school years), and zero otherwise; \mathbf{X} is a vector of time-varying district-specific characteristics, including the percent of district i ’s students in year t receiving free/reduced-price lunch, the percent of minority students (i.e., African American or Hispanic), the percent of English language learners, the percent of special education/disabled students, serious incidents (per-capita), and the proportion of students academically proficient on the math portion of the state’s accountability exam.⁷ The variables θ_i and λ_t are district and year fixed effects, respectively; $\gamma_i t$ is the set of district-specific linear time trends, which capture any pre-reform outcome trend differences, as seen in Figure 1; and ε_{it} is a random error term. We later discuss the extent to which our results are sensitive to the inclusion of these district-specific

⁷ Because math and reading achievement are highly correlated ($r=.93$), we include only math achievement as a predictor.

trends. We cluster the standard errors at the district level. The coefficient β_1 is the DD estimate of the effect of Philadelphia’s discipline policy reform on suspension usage.⁸

A primary concern in the context of the panel DD approach relates to time-varying district-specific (unobserved) heterogeneity that, if related to both the timing of discipline policy reform and the outcomes of interest, would introduce bias into the DD estimates. Indeed, as Angrist and Pischke (2009) note, the inclusion of a district-specific time trend “allows treatment and control (districts) to follow different trends in a limited but potentially revealing way. It’s heartening to find that the estimated effects of interest are unchanged by the inclusion of these trends, and discouraging otherwise” (pgs. 238-239). For each model, we estimate effects both with and without district-specific time trends. Given that the outcome trends vary across Philadelphia and the 482 comparison districts (see Figure 1 and Figure A2), we prefer results that control for district-specific linear time trends.

⁸ Though the DD approach allows us to account for district-specific time trends and time invariant district effects, each comparison district, no matter how different from Philadelphia on observable dimensions, contributes equally to the DD estimate of the effect of the policy reform. A synthetic control group method is an approach which empirically matches districts that more closely resemble the demographic, academic and pre-reform outcome trends found in Philadelphia, and weights those districts (based on the extent to which they resemble pre-reform Philadelphia) in the construction of a control group of comparison districts. The synthetic control group approach generates a comparison group from the pool of Pennsylvania districts based on existing pre-reform outcome trends (Abadie, Diamond, and Hainmueller 2010; Abadie and Gardeazabal 2003; Bassok, Fitzpatrick, and Loeb 2014; Eren and Ozbeklik 2016; Grindal, West, Willett, and Yoshikawa 2015; Rubin 2001; Rudolph, Stuart, Vernick, and Webster 2015). The comparison group (*synthetic Philadelphia*) in the synthetic control group approach is a weighted combination of all comparison districts based on the similarity of each district in the pre-reform period to Philadelphia, the treated unit. This synthetic control group approach compares suspensions (and other school and student outcomes) in Philadelphia before and after the policy change to the suspensions (and other outcomes) we would expect to occur in the absence of the policy change. The guiding assumption is that districts most similar to the treatment district provide the best possible counterfactual for Philadelphia in the absence of discipline policy reform. In earlier versions of this paper, we estimated the impact of the policy reform using both DD and synthetic control group approaches. The synthetic control group approach accounts for many ways in which a rural Pennsylvania district differs from urban Philadelphia, including, for example, student characteristics and the prevalence of out-of-school suspensions. However, in practice, Philadelphia differs dramatically from most Pennsylvania districts, and only a handful of the 482 comparison districts (between five and seven, depending on the outcome measure) received nonzero weight. Further, the synthetic control group approach relies on a stringent requirement that the pre-reform outcome trends of actual and synthetic Philadelphia match, a condition that the DD approach does not require. Violations to this condition may bias estimates of the effect of discipline policy reform in Philadelphia generated by the synthetic control group approach. In our synthetic control models, we were not able to generate close matches in the pre-reform OSS trends. In the DD models, the inclusion of district-specific linear time trends explicitly accounts for any pre-reform outcome trend differences across Philadelphia and each set of comparison districts. Therefore, we concluded that the results of the synthetic control group models were not as reliable as those generated from the difference-in-difference models with district-specific linear time trends, which we present here (results based on the synthetic control group approach are available from the authors upon request).

Results

Effect of the Policy Change on OSS

We find that the introduction of discipline policy reform in Philadelphia had little overall effect on suspension usage in the post-policy period. While estimates indicate that discipline policy reform increased total OSS in Philadelphia by 1.6 suspensions per 100 students over the 2012-13 through 2014-15 school years, there is variation by post-policy year (Table 2). Specifically, there was no change in overall suspension usage in the first two years following the policy change, and then an increase in suspension usage in the third post-policy year. Following Angrist and Pischke (2009), we estimate models both with and without district-specific time trends. We find that the results are highly sensitive to the inclusion of these trends, with the direction of the effects changing once the district time trends are appropriately controlled for, reinforcing the importance of conditioning on district specific time trends in light of variation in pre-reform outcome trends.

Given that Philadelphia's revised code of conduct centered on reducing the number of suspensions for lower-level, non-violent infractions without eliminating suspensions for more serious offenses, we would expect any reduction in OSS to be concentrated among conduct suspensions. In fact, we find that, on average, OSS for conduct infractions remained unchanged across the three post-policy years (Table 2). This null finding, however, obscures year-specific variation. Specifically, in the first post-policy year (i.e., 2012-13), conduct suspensions declined by 1 conduct suspension per 100 students. In the second post policy year, the coefficient is negative but the effect does not reach statistical significance. In the third post-policy year, we find that conduct suspensions increased by 4 per 100 students. These models show evidence of initial, modest declines in conduct suspensions following the reform, which did not persist past the first post-policy year.

Philadelphia's discipline policy reform also provided school leaders with greater discretion in handling more serious student misconduct. In addition, if discipline policy reform had broader effects on school climate and students' perceptions of disciplinary fairness, we might observe changes in the use of more serious, nonconduct OSS (Table 2). We find that nonconduct OSS in Philadelphia actually increased following the policy change, by a larger margin than the change in conduct OSS. Specifically, the DD results indicate that nonconduct OSS increased by 0.014, or 1.4 more suspensions per 100 students over the post-policy period. One might be concerned that the implementation of Philadelphia's revised code of conduct had an effect on OSS in neighboring school districts. As Rudolph et al. (2015) note, one approach to address this assumption of no interference between the treatment unit (Philadelphia) and comparison districts is to restrict the analysis to districts that do not geographically border Philadelphia. When we restrict the sample of comparison districts by excluding those districts that border Philadelphia, we estimate average post-policy effects that are nearly identical to the main results.⁹

⁹ We drop districts in Bucks, Delaware and Montgomery counties, which technically border Philadelphia. Results are available from the authors upon request.

Table 2. Effect of Discipline Policy Reform on OSS, by OSS type

	All OSS			Conduct OSS			Non-conduct OSS		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Philadelphia*PostPolicy	-0.044*** (0.008)	0.016*** (0.005)		-0.031*** (0.008)	0.001 (0.005)		-0.015*** (0.002)	0.014*** (0.001)	
Philadelphia*2012-13			0.006 (0.005)			-0.010** (0.004)			0.014*** (0.001)
Philadelphia*2013-14			0.005 (0.006)			-0.003 (0.006)			0.007*** (0.001)
Philadelphia*2014-15			0.067*** (0.008)			0.040*** (0.008)			0.026*** (0.002)
District-Specific Trends	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Districts	483	483	483	483	483	483	483	483	483
Observations	4,830	4,830	4,830	4,830	4,830	4,830	4,830	4,830	4,830
R-squared	0.707	0.791	0.791	0.666	0.756	0.756	0.700	0.784	0.784

Note. Each column represents a separate regression. Coefficients are from difference-in-differences models and are reported with standard errors (clustered at the district level) in parentheses. *Philadelphia*PostPolicy* is the average effect in the post-policy period (i.e., 2012-13 through 2014-15). Models include district fixed effects, year fixed effects and controls for the following district characteristics: the percent of students receiving free or reduced-price lunch, the percent of minority (i.e., African American or Hispanic) students, the percent of English language learners, the percent of special education/disabled students, serious incidents (per-capita), and district math achievement (the percent of district students academically proficient on the PSSA math exam). Models in columns 2, 3, 5, 6, 8 and 9 include district-specific linear time trends. Coefficients statistically significant at the *10%, **5%, and ***1% levels.

Exploring Changes in Specific Infractions Subject to Discipline Policy Reform

Not only did the district's policy reform formally prohibit OSS for two low-level, conduct infractions – failure to follow classroom rules and profane language – but the reform also changed the most punitive disciplinary response – from expulsion to OSS – for more serious student misconduct (see Table A1). Given the estimated decline in conduct OSS and corresponding increase in nonconduct OSS in the first two post-policy years, we examine the patterns of OSS for each of the infractions targeted by the district's policy reform. To do so, we rely on detailed student-level information on infractions and suspensions for the 2011-12 through 2013-14 school years. Specifically, we observe the infraction type, code of conduct rule number and offense type associated with each observed OSS.¹⁰

In 2011-12, 24.9 percent of all suspensions given to grade 3-12 students in Philadelphia were for either failure to follow classroom rules or profane/obscene language or gestures. By 2012-13, the first year of the policy reform, 15.1 percent of all suspensions were for these two infractions; in 2013-14, 12.1 percent of all suspensions were for these two infractions (see Figure 2).¹¹ In per-capita terms, there were approximately 7 OSS per 100 students in 2011-12; this number declined by more than half in the two post-policy years, to 3.6 OSS and 3.2 OSS per 100 students in 2012-13 and 2013-14, respectively (see Figure 2).

Though we find substantive declines in OSS among the conduct infractions targeted by the policy reform,¹² we also observe meaningful increases in OSS among more serious, nonconduct infractions. In 2011-12, 15.9 percent of all OSS in 2011-12 were for one of the ten more serious infractions for which the district's policy reform changed the most punitive disciplinary response from expulsion to OSS. The OSS rate for more serious misconduct increased to 36.1 percent in 2012-13 and 31.7 percent in 2013-14. Further, the per-capita number of OSS for more serious student misconduct doubled in the post-policy period, from 4.4 OSS per 100 students in 2011-12, to 8.6 OSS per 100 students in 2012-13 (and 8.3 OSS per 100 students in 2013-14). These post-policy increases were driven by increases in OSS for three specific infractions: harassment/bullying/cyber-bullying/intimidation; threatening students/staff with aggravated assault; and simple assault (documented unprovoked attack by one student on another).

The post-policy increase in nonconduct OSS could be the result of many factors. First, schools may have complied with the policy reform which required a shift in disciplinary consequences away from expulsion to OSS as the most punitive response allowable. Second, a change in student behavior could have led to more nonconduct OSS. This likely explains some of the observed increase in nonconduct OSS given that we estimate an increase in serious incidents following the policy reform (see Table 3). Third, schools may have strategically reclassified student misconduct away from infractions for which OSS was formally prohibited (i.e., failure to follow classroom rules and profane language) to infraction types subject to OSS (i.e., bullying).

¹⁰ The authors thank the School District of Philadelphia for providing the student-level data to conduct this analysis.

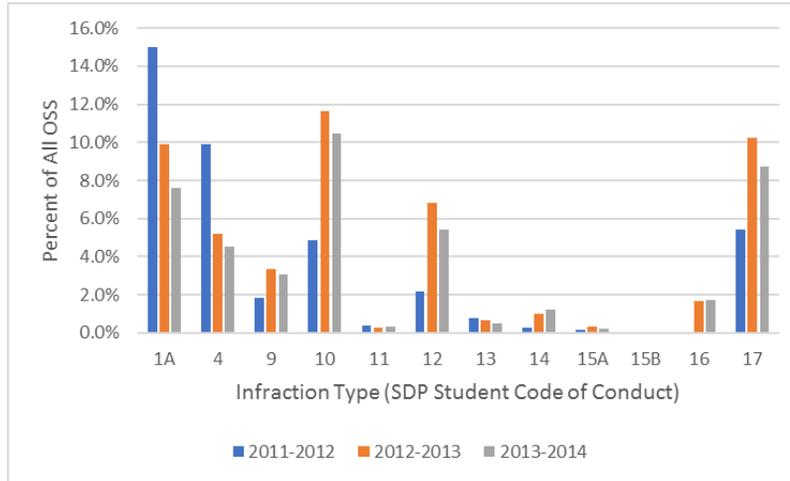
¹¹ A total of 27903, 23584, and 26838 suspensions were given to grade 3-12 students in the 2011-12, 2012-13, 2013-14 school years, respectively.

¹² Notably, the use of OSS for failure to follow classroom rules or profane/obscene language or gestures was not eliminated in the post-policy period. See Steinberg & Lacoë (2017) for a detailed analysis of school-level implementation of the district's policy reform, which finds significant variation in school-level compliance toward eliminating OSS for these conduct infractions in the post-policy period.

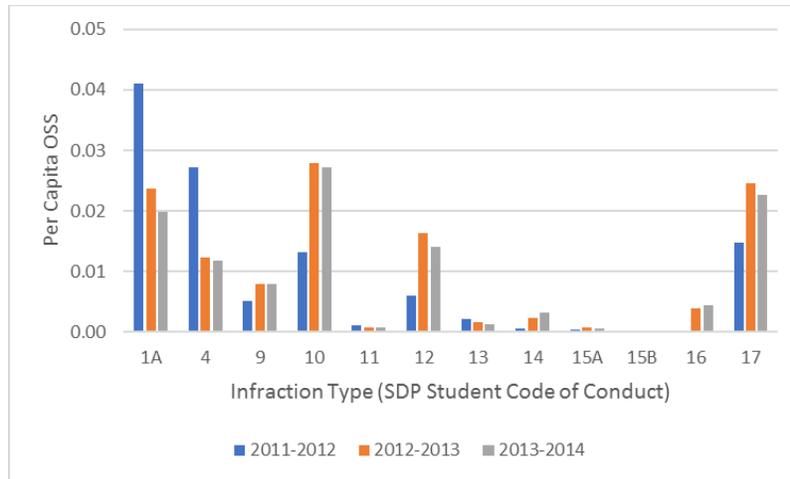
While we are unable to attribute changes in nonconduct OSS to these (and potentially other) factors, these patterns suggest that district-level policy reforms must be cognizant of the potential off-setting increase in suspensions for more serious infractions if the aim is to reduce the overall use of OSS districtwide.

Figure 2. OSS by Infraction Type Targeted by District Policy Reform

Panel A: Percent of All OSS



Panel B: Per Capita OSS



Note. See Table A1 for a description of each of the twelve infraction types (as defined by the SDP student code of conduct) targeted by the district’s policy reform. Sample includes all grade 3-12 students in traditional public schools in Philadelphia. In Panel A, figure shows the percent of all OSS by infraction type; a total of 27903, 23584, and 26838 suspensions were given to grade 3-12 students in the 2011-12, 2012-13, 2013-14 school years, respectively. In Panel B, figure shows the per-capita number of OSS by infraction type; total grade 3-12 enrollment in Philadelphia was 101844, 98824, and 103158 students in the 2011-12, 2012-13, 2013-14 school years, respectively. In 2011-12, infraction type 16 (Mutual fight (w/ documented serious bodily injury) was not available in the data; infraction types 9 (Destruction and/or theft of property (less than \$1,000)) and 13 (Destruction and/or theft of property (totaling \$1,000 or more)) were for \$150 in 2011-12 and \$1000 in 2012-13 and 2013-14.

Effects on Student Misconduct, Attendance and Achievement

The policy change in Philadelphia decreased suspensions for nonviolent conduct offenses, which were the target of the district’s code of conduct reform, in the first post-policy year. In addition to changing the statutory penalties for specific infractions, the discipline policy reform provided school leaders in Philadelphia with greater discretion in handling more serious student misconduct, such as destruction or theft of property, bullying, fighting, and alcohol and drug possession. As a result, discipline policy reform may have also affected other aspects of school climate – such as student behavior and engagement – and student achievement. We extend the DD analysis to explore these potential policy-related effects. We note that these estimates are reduced form in nature, and should be interpreted as intent-to-treat (ITT) effects of the policy reform rather than the effect of changes in OSS on school and student outcomes.

First, we find that the policy reform is associated with declines in student behavior. Evidence from the DD approach suggests that serious incidents of student misconduct increased in Philadelphia, on the order of 1.3 additional serious incidents per 100 students, over the post-policy period (Table 3, Column 2). The increase in serious incidents is consistent in sign and magnitude across all three post-policy years. This provides evidence of a relationship between the change in discipline policy and worsening student behavior, even after controlling for changes in OSS. The results suggest that the policy change, which included greater principal discretion in handling more serious student misconduct, contributed to increases in school violence and disruption.

Table 3. Effect of Discipline Policy Reform on Student Misconduct and Truancy

Outcome Variable	Serious Incidents			Truancy Rate		
	(1)	(2)	(3)	(4)	(5)	(6)
Philadelphia*PostPolicy	-0.016*** (0.004)	0.013*** (0.002)		0.019* (0.011)	0.080*** (0.006)	
Philadelphia*2012-13			0.012*** (0.002)			0.039*** (0.005)
Philadelphia*2013-14			0.014*** (0.003)			0.099*** (0.007)
Philadelphia*2014-15			0.012*** (0.004)			0.166*** (0.010)
District-Specific Trends	No	Yes	Yes	No	Yes	Yes
Districts	483	483	483	483	483	483
Observations	4,830	4,830	4,830	4,830	4,830	4,830
R-squared	0.483	0.647	0.647	0.524	0.671	0.671

Note. Each column represents a separate regression. Coefficients are from difference-in-differences models and are reported with standard errors (clustered at the district level) in parentheses. *Philadelphia*PostPolicy* is the average effect in the post-policy period (i.e., 2012-13 through 2014-15). Models include district fixed effects and year fixed effects; Models in columns 2,3, 5 and 6 include district-specific linear time trends. All models include controls for the following district characteristics: the percent of students receiving free or reduced-price lunch, the percent of minority

(i.e., African American or Hispanic) students, the percent of English language learners, the percent of special education/disabled students and Total OSS (per-capita), and math achievement (the percent of district students academically proficient on the PSSA math exam). Coefficients statistically significant at the *10%, **5%, and ***1% levels.

Second, the policy change decreased student engagement, as measured by truancy, or persistent unexcused absence from school. We find that the policy change is associated with an increase in the truancy rate over the post-policy period of 8.0 percentage points (Table 3, column 5). Increases in the truancy rate grew larger in each subsequent post-policy year. Though one might expect that greater inclusion of students who previously would have been suspended could lead to a decrease in truancy, we find that disengagement among students outweighs any potential improvements in attendance generated by the policy.

Third, we find evidence that math and ELA proficiency at the district level declined following the policy change (Table 4). Relative to other districts during the same period, math and ELA achievement declined in Philadelphia in the post-policy years. Overall, the achievement models suggest significant academic costs associated with Philadelphia's policy change.

Table 4. Effect of Discipline Policy Reform on Student Achievement

Outcome Variable	Math Achievement			ELA Achievement		
	(1)	(2)	(3)	(4)	(5)	(6)
Philadelphia*PostPolicy	-2.35** (0.990)	-5.28*** (0.439)		-0.64 (0.610)	-4.26*** (0.298)	
Philadelphia*2012-13			-5.66*** (0.398)			-3.46*** (0.272)
Philadelphia*2013-14			-5.83*** (0.511)			-4.24*** (0.349)
Philadelphia*2014-15			-3.05*** (0.626)			-6.72*** (0.443)
District-Specific Trends	No	Yes	Yes	No	Yes	Yes
Districts	483	483	483	483	483	483
Observations	4,830	4,830	4,830	4,830	4,830	4,830
R-squared	0.927	0.959	0.959	0.924	0.955	0.955

Note. Each column represents a separate regression. Coefficients are from difference-in-differences models and are reported with standard errors (clustered at the district level) in parentheses. *Philadelphia*PostPolicy* is the average effect in the post-policy period (i.e., 2012-13 through 2014-15). Models include district fixed effects and year fixed effects; Models in columns 2, 3, 5 and 6 include district-specific linear time trends. All models include controls for the following district characteristics: the percent of students receiving free or reduced-price lunch, the percent of minority (i.e., African American or Hispanic) students, the percent of English language learners, the percent of special education/disabled students and Total OSS (per-capita). Coefficients statistically significant at the *10%, **5%, and ***1% levels.

Conclusion

This paper provides empirical evidence on the efficacy of a district-level policy change in Philadelphia aimed at reforming zero-tolerance approaches to discipline and reducing out-of-school suspensions. Though nearly a quarter of the nation's largest school districts have recently amended their student codes of conduct in an effort to roll back zero-tolerance responses to student misbehavior, relatively little is known about whether such policy reforms in fact reduce suspension usage. We find evidence that the policy change resulted in very modest declines in conduct OSS immediately following the change, but these decreases in conduct OSS did not persist. These results suggest that the policy reform was not successful in achieving its aim of reducing the use of suspensions for low-level student misconduct in the longer term. We also investigate the broader relationship between discipline policy reform and school climate and student achievement, and identify increases in serious behavioral incidents, declines in student engagement (via increases in the truancy rate), and declines in achievement. These findings have important implications for policy and practice, as districts and states across the nation consider similar reforms aimed at reducing suspensions.

First, the discipline policy change in Philadelphia initially reduced suspensions for conduct infractions targeted by the reform, but those decreases did not persist. Given that principals have relative autonomy from district policymakers to implement policy reforms (Lipsky 2010), implementation heterogeneity at the school level may explain why low-level conduct suspensions did not decline in each of the post-policy years. Notably, Steinberg & Lacoë (2017) document significant variation in policy implementation across Philadelphia schools which may have limited the policy's effect on suspension usage. Indeed, evidence indicates that schools serving more economically and educationally disadvantaged students in Philadelphia – those with higher pre-reform OSS rates and lower academic achievement levels – *increased* their use of conduct suspensions in the post-reform period (Steinberg & Lacoë, 2017). This underscores the fact that the efficacy of district-level policy reform will be constrained in important ways by differences in school-level capacity to reduce the use of OSS.

Second, reducing the statutory penalties for more serious student misconduct— from expulsion to suspension – may lead to a worsening of school climate. This is in light of the fact that, compared to other districts in Pennsylvania, Philadelphia schools experienced increases in serious incidents of student misconduct as well as increases in suspensions for nonconduct (e.g. more serious) infractions in the years following the policy change. Further, we find that the policy reform increased the truancy rate among students in Philadelphia while also leading to declines in student achievement. To minimize the potential for such unintended consequences, any changes to a district's code of conduct that are jointly designed to reduce suspensions and expulsions should be coupled with additional supports for schools that face disciplinary challenges. This is in light of the fact that Philadelphia provided few (if any) additional resources or programmatic supports to schools to implement the districtwide reform.

Finally, this paper contributes important evidence on the impact of a discipline policy reform in one district, but the results raise several important questions about the implementation of district-level discipline policy reform that are more difficult to answer with administrative data. For example, did schools implement any other discipline practices with the introduction of the district's policy reform (i.e., PBIS, restorative justice, etc.)? If so, did schools have sufficient

resources and support to fully implement these practices? Did schools substitute away from OSS to other types of discipline in response to low-level conduct infractions, such as in-school suspension, in the first post-reform year? Did any institutional factors, such as resource levels, staff capacity, or principal leadership, limit schools' efforts to fully implement the district's policy reform? School-level differences in resource capacity are important to consider, given evidence that schools serving different student populations varied in their implementation of the policy reform (Steinberg & Laco, 2017). As districts across the nation enact discipline policy reforms, this empirical investigation of Philadelphia's reform effort should serve as a guide to better understand the implications of such policy reforms on the use of suspensions and the potential consequences for school climate and student achievement.

References

- Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, 105(490), 493–505.
- Abadie, A., & Gardeazaball, J. (2003). The economic costs of conflict: A case study of the Basque country. *American Economic Review*, 93(1), 112–132.
- Allensworth, E., Ponisciak, S., & Mazzeo, C. (2009). The schools teachers leave: Teacher mobility in Chicago public schools. Chicago: Consortium on Chicago School Research.
- American Academy of Pediatrics. (2003). Out-of-school suspensions and expulsion. *Pediatrics*, 112(5), 1206–1209.
- American Association of School Superintendents. (2013). School discipline consensus project survey of school system leaders. Retrieved from <http://blogs.edweek.org/edweek/rulesforengagement/AASA%20School%20Climate%20Survey.pdf>
- American Psychological Association Zero Tolerance Task Force. (2008). Are zero tolerance policies effective in schools? An evidentiary review and recommendations. *American Psychologist*, 63(9), 852–862.
- Angrist, J.D., & Pischke, J. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton: Princeton University Press.
- Anyon, Y., Gregory, A., Stone, S., Farrar, J., Jenson, J. M., McQueen, J., Downing, B., Greer, E., & Simmons, J. (2016). Restorative interventions and school discipline sanctions in a large urban school district. *American Educational Research Journal*, 53(6), 1663-1697.
- Arum, R. (2003). *Judging school discipline: The crisis of moral authority*. Cambridge, MA: Harvard University Press.
- Baker, A. (2012, August 28). New code aims to ease suspensions of students. *The New York Times*. Retrieved from <http://www.nytimes.com/2012/08/29/education/new-york-education-department-revises-student-disciplinary-code.html>
- Baker, M. L., Sigmon, J. N., & Nugent, M. E. (2001). Truancy reduction: Keeping students in school. Office of Justice Programs, U.S. Department of Justice.
- Bassok, D., Fitzpatrick, M., & Loeb, S. (2014). Does state preschool crowd-out private provision? The impact of universal preschool on the childcare sector in Oklahoma and Georgia. *Journal of Urban Economics*, 83, 18-33.
- 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.
- Bowen, N.K., Bowen, G.L., & Ware, W.B. (2002). Neighborhood social disorganization, families, and the educational behavior of adolescents. *Journal of Adolescent Research*, 17(5), 468–90.
- Bradshaw, C.P., Mitchell, M.M., & Leaf, P.J. (2010). Examining the effects of schoolwide positive behavioral interventions and supports on student outcomes: Results from a randomized controlled effectiveness trial in elementary schools. *Journal of Positive Behavior Interventions*, 12(3), 133-148.

- Bradshaw, C.P., Sawyer, A.L., & O'Brennan, L.M. (2009). A social disorganization perspective on bullying-related attitudes and behaviors: The influence of school context. *American Journal of Community Psychology*, 43, 204–220.
- Bryk, A.S., Sebring P.B., Allensworth, E., Luppescu, S., & Easton, J.Q. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago: The University of Chicago Press.
- Carrell, S.E., Hoekstra, M., & Kuka, E. (2016, February). The long-run effects of disruptive peers. *NBER Working Paper* No. 22042.
- Courtemanche, C.J., & Zapata, D. (2014). Does universal coverage improve health? The Massachusetts experience. *Journal of Policy Analysis and Management*, 33(1), 36–69.
- Curran, F. C. (2016). Estimating the Effect of State Zero Tolerance Laws on Exclusionary Discipline, Racial Discipline Gaps, and Student Behavior. *Educational Evaluation and Policy Analysis*, 0162373716652728.
- Eckholm, E. (2013, April 12). With police in schools, more children in court. *The New York Times*. Retrieved from <http://www.nytimes.com/2013/04/12/education/with-police-in-schools-more-children-in-court.html?pagewanted=all>
- Eren, O., & Ozbeklik, S. (2016). What do right-to-work laws do? Evidence from a synthetic control method analysis. *Journal of Policy Analysis and Management*, 35(1), 173–194.
- Fabelo, T., Thompson, M.D., Plotkin, M., Carmichael, D., Marchbanks, M.P., & Booth, E.A. (2011). *Breaking schools' rules: A statewide study of how school discipline relates to students' success and juvenile justice involvement*. Justice Center, Council of State Governments.
- Felson, R.B., Liska, A.E., South, S.J., & McNulty, T. (1994). The subculture of violence and delinquency: Individual vs. school context effects. *Social Forces*, 73(1), 155–173.
- Freeman, J., Simonsen, B., McCoach, D. B., Sugai, G., Lombardi, A., & Horner, R. (2016). Relationship between school-wide positive behavior interventions and supports and academic, attendance, and behavior outcomes in high schools. *Journal of Positive Behavior Interventions*, 18(1), 41-51.
- Gottfried, M. A. (2014). The achievement effects of tardy classmates: Evidence in urban elementary schools. *School Effectiveness and School Improvement*, 25(1), 3-28.
- Gregory, A., Cornell, D., & Fan, X. (2011). The relationship of school structure and support to suspension rates for Black and White high school students. *American Educational Research Journal*, 48(4), 904-934.
- Grindal, T., West, M.R., Willett, J.B., & Yoshikawa, H. (2015). The impact of home-based child care provider unionization on the cost, type, and availability of subsidized child care in Illinois. *Journal of Policy Analysis and Management*, 34(4), 853–880.
- Hanson, A.L. (2005). Have zero tolerance school discipline policies turned into a nightmare? The American dream's promise of equal educational opportunity grounded in *Brown v. Board of Education*. *UC Davis Journal of Juvenile Law & Policy*, 9(2), 289–379.
- Heckman, J.J., & Hotz, V.J. (1989). Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training. *Journal of the American Statistical Association*, 84(408), 862–874.
- Hoffman, S. (2014). Zero benefit: Estimating the effect of zero tolerance discipline policies on racial disparities in school discipline. *Educational Policy*, 28(1), 69–95.

- Imberman, S.A., Kugler, A.D., & Sacerdote, B.I. (2012). Katrina's children: Evidence on the structure of peer effects from hurricane evacuees. *American Economic Review*, *102*(5), 2048-2082.
- Kaul, A., Klobner, S., Pfeifer, G., & Schieler, M. (2015). Synthetic control methods: Never use all pre-intervention outcomes as economic predictors. Working Paper.
- Kraft, M., Marinell, W.H., & Yee, D.S. (2016). School organizational contexts, teacher turnover, and student achievement: Evidence from panel data. *American Educational Research Journal*, *53*(5), 1411-1449.
- Kupchik, A., & Ellis, N. (2008). School discipline and security fair for all students? *Youth & Society*, *39*(4), 549–574.
- Lacoe, J. (2015). Unequally safe: The race gap in school safety. *Youth Violence and Juvenile Justice*, *13*(2), 143–168.
- Lacoe, J. (2016). Too scared to learn? The academic consequences of feeling unsafe in the classroom. *Urban Education*.
- Lacoe, J., & Steinberg, M.P. (2017). The impact of out-of-school suspensions on student outcomes. Working Paper.
- Lipsky, M. (2010). *Street-level bureaucracy: Dilemmas of the individual in public services*. New York: Russell Sage Foundation.
- Losen, D.J., & Gillespie, J. (2012). Opportunities suspended: The disparate impact of disciplinary exclusion from school. Los Angeles, CA: The Civil Rights Project/Proyecto Derechos Civiles. Retrieved from <http://civilrightsproject.ucla.edu/resources/projects/center-for-civil-rights-remedies/school-to-prison-folder/federal-reports/upcoming-crr-research>
- Macmillan, R., & Hagan, J. (2004). Violence in the transition to adulthood: Adolescent victimization, education, and socioeconomic attainment in later life. *Journal of Research on Adolescence*, *14*(2), 127–158.
- Mader, N., Sartain, L., & Steinberg, M. (2016). When suspensions are shorter: The effects on school climate and student learning. Working Paper.
- Mendez, L.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.
- Mezzacappa, D. (2012, August 17). SRC adopts revised student code of conduct. *The Notebook*. Retrieved from <http://thenotebook.org/blog/125075/src-adopts-revised-student-code-conduct>
- Mitchell, M. M., & Bradshaw, C. P. (2013). Examining classroom influences on student perceptions of school climate: The role of classroom management and exclusionary discipline strategies. *Journal of School Psychology*, *51*(5), 599-610.
- Neiman, S. (2011). Crime, violence, discipline, and safety in U.S. public schools: Findings from the School Survey on Crime and Safety: 2009-10. NCES 2011-320. Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- New York City School-Justice Partnership Task Force. (2013). Keeping kids in school and out of court. Report for the New York State Permanent Commission on Justice for Children. Retrieved from <http://www.courts.state.ny.us/IP/justiceforchildren/PDF/NYC-School-JusticeTaskForceReportAndRecommendations.pdf>

- Paslay, C. (2012, August 17). The day discipline died in Philadelphia [Web log post]. Retrieved from <http://chalkandtalk.wordpress.com/2012/08/17/the-day-discipline-died-in-philadelphia-public-schools/>
- Payne, A.A., Gottfredson, D.C., & Gottfredson, G.D. (2003). Schools as communities: The relationships among communal school organization, student bonding, and school disorder. *Criminology*, 41(3), 749–777.
- Peguero, A. A., & Bracy, N. L. (2015). School order, justice, and education: climate, discipline practices, and dropping out. *Journal of research on adolescence*, 25(3), 412-426.
- Pennsylvania Information Management System, User Manual (Volume 1), 2014-15; page 79.
- Osher, D., Bear, G.G., Sprague, J.R., & Doyle, W. (2010). How can we improve school discipline? *Educational Researcher*, 39(1), 48–58.
- Raffaele Mendez, L. M. (2003). Predictors of suspension and negative school outcomes: A longitudinal investigation. *New directions for youth development*, 2003(99), 17-33.
- Rausch, M.K., & Skiba, R.J. (2006). The academic cost of discipline: The relationship between suspension/expulsion and school achievement. Bloomington, IN: Center for Evaluation and Education Policy, Indiana University.
- Rich, M. (2012, August 7). Suspensions are higher for disabled students, federal data indicate. *The New York Times*. Retrieved from http://www.nytimes.com/2012/08/08/education/analysis-examines-disabled-students-suspensions.html?_r=1&ref=education
- Rubin, D.B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 1, 169–188.
- Rudolph, K.E., Stuart, E.A., Vernick, J.S., & Webster, D.W. (2015). Association between Connecticut’s permit-to-purchase handgun law and homicides. *American Journal of Public Health*, 105(8), 49–54.
- Schreck, C. J., & Miller, J.M. (2003). Sources of fear of crime at school: What is the relative contribution of disorder, individual characteristics, and school security? *Journal of School Violence*, 2(4), 57–79.
- Shah, N. (2012, August 22). Groups ask districts to stop using out-of-school suspensions. *Education Week*. Retrieved from http://blogs.edweek.org/edweek/District_Dossier/2012/08/groups_ask_districts_to_stop_u_h_tml
- Shah, N. (2013, April 8). New analysis bolsters case against suspension, researchers say. *Education Week*. Retrieved from http://blogs.edweek.org/edweek/rulesforengagement/2013/04/new_data_analysis_bolsters_case_against_suspension_researchers_say.html
- Skiba, R.J. (2008). Are zero tolerance policies effective in the schools? An evidentiary review and recommendations. *American Psychologist*, 63(9), 852–862.
- 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.
- Smith, D.L., & Smith, B.J. (2006). Perceptions of violence: The views of teachers who left urban schools. *The High School Journal*, 89(3), 34–42.

- Snyder, S. (2012, August 13). School summit focuses on violence prevention. *The Philadelphia Inquirer*. Retrieved from http://articles.philly.com/2012-08-13/news/33183381_1_school-safety-philadelphia-school-district-principals-safer-schools
- Steinberg, M., Allensworth, E., & Johnson, D.W. (2011). Student and teacher safety in Chicago public schools: The roles of community context and school social organization. Chicago: Consortium on Chicago School Research.
- Steinberg, M., & Lacoë, J. (2017). What do we know about school discipline reform? Assessing the alternatives to suspensions and expulsions. *Education Next*, 17(1).
- Taube, D. (2013, May 3). Discipline do-over: San Diego Unified is rethinking its zero-tolerance policy. *San Diego CityBeat*. Retrieved from <http://www.sdcitybeat.com/sandiego/blog-1203-discipline-do-over.html>
- Watanabe, T. (2013, May 14). L.A. 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. (2013, May 15). L.A. schools will no longer suspend a student for being defiant. *Los Angeles Times*. Retrieved from <http://www.latimes.com/local/lanow/la-me-ln-laUSD-suspensions-20130515,0,442761.story>
- Welsh, W.N., Greene, J.R., & Jenkins, P.H. (1999). School disorder: The influence of individual, institutional, and community factors. *Criminology*, 37(1), 73–115.
- Yee, V. (2013, May 6). Protesters target Weiner on school discipline stance. *The New York Times*. Retrieved from <http://cityroom.blogs.nytimes.com/2013/05/06/protesters-target-weiner-on-school-discipline-stance/>

Appendix

Table A1. SDP Code of Student Conduct: Changes to Most Punitive Disciplinary Responses

Rule	Offense	Most punitive disciplinary response	
		2011–12 (pre-policy)	2012–13 (post-policy)
1A	Failure to follow classroom rules/disruption	Out-of-school suspension (1–3 days)	In-school intervention
4	Profane/obscene language or gestures		
9	Destruction and/or theft of property (less than \$1,000)	Expulsion	Out-of-school suspension; lateral transfer
10	Harassment/bullying/cyber-bullying/intimidation	Expulsion	Out-of-school suspension; lateral transfer; disciplinary school assignment
11	Sexual acts (consensual)	Expulsion	Out-of-school suspension; lateral transfer
12	Threatening students/staff with aggravated assault		
13	Destruction and/or theft of property (totaling \$1,000 or more)		
14	Breaking and entering school property		Out-of-school suspension; lateral transfer; disciplinary school assignment
15A	Robbery	Expulsion	
15B	Extortion		
16	Mutual fight (w/ documented serious bodily injury)		
17	Simple assault (documented unprovoked attack by one student on another)		

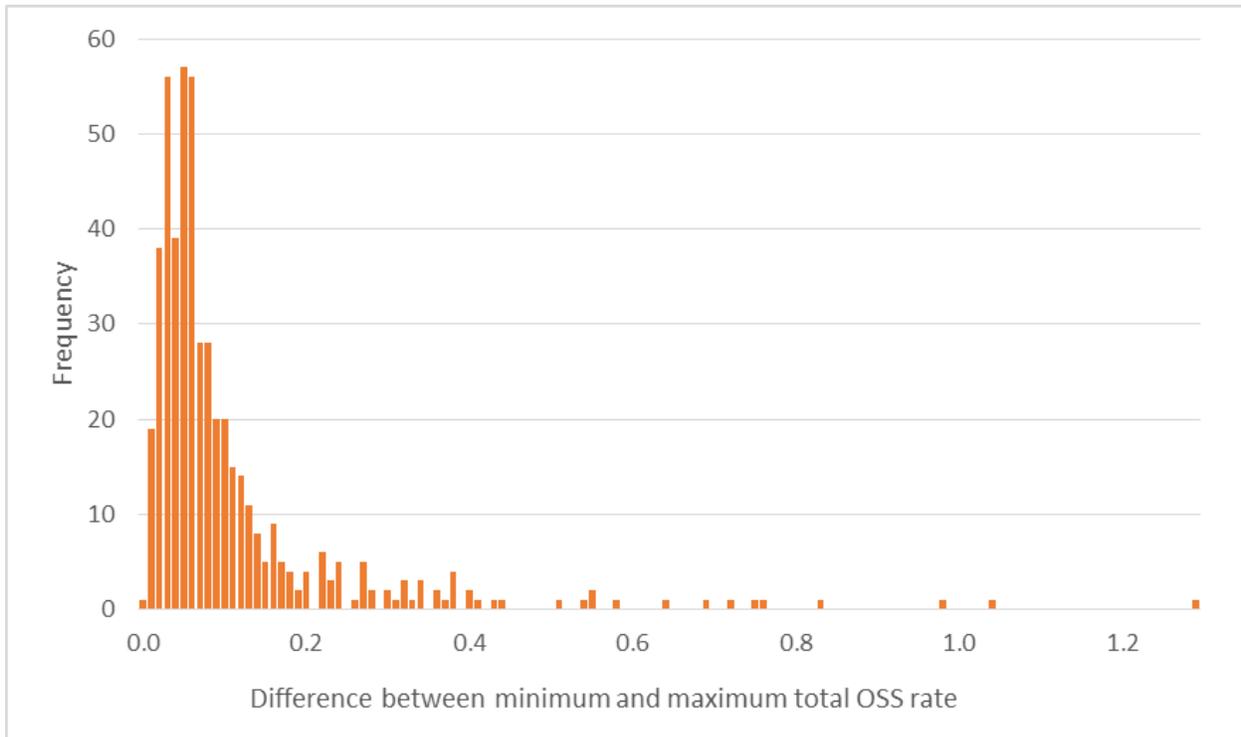
Note. Authors' summary from The School District of Philadelphia 2011–2012 Code of Student Conduct and The School District of Philadelphia 2012–13 Code of Student Conduct. Rule number based on the 2012–2013 Code of Student Conduct. *In-school intervention* includes in-school suspension. *Lateral transfer* refers to a transfer to another school setting. *Disciplinary school assignment* refers to a student's transfer to a disciplinary school.

Table A2. Crosswalk between Pennsylvania Department of Education (PDE) Out-of-School Suspension (OSS) Category and School District of Philadelphia (SDP) Student Misconduct Incident

OSS Reporting Category (PDE)	Student Misconduct Incident (SDP)
Conduct	Alteration of grade reporting/excuses/school documents; Dress code violation; Failure to carry hall-pass/appropriate ID; Failure to follow classroom rules/disruption; Failure to participate in class/unpreparedness; Forgery of administrator, teacher, or parent's/guardian's signature; Inappropriate use of electronic devices; Possession of beepers/pagers/cell phones/other devices; Possession of other inappropriate personal items; Profane/obscene language or gestures; Public display of affection/inappropriate touching; and Truancy/excessive tardiness/cutting class
Drugs and Alcohol	Possession of alcohol or drugs - intent to distribute; Possession of alcohol or drugs – intent to use
Violence	Aggravated assault (documented serious bodily injury); Assault of school personnel; Breaking and entering school property; Destruction and/or theft of property (less than \$1,000); Destruction and/or theft of property (totaling \$1,000 or more), Extortion; Fighting (two students engaged in mutual combat); Harassment/bullying/cyber-bullying/intimidation; Instigation or participation in group assaults; Mutual fight (w/ documented serious bodily injury); Robbery; Sexual acts (consensual); Sexual acts (non-consensual); Simple assault (documented unprovoked attack by one student on another); and Threatening students/staff with aggravated assault
Weapon	Possession of a weapon; Possession or use of fireworks/incendiary devices/explosives

Note. Conduct infractions are reported to PDE in aggregate and not by specific incident type. SDP's Student Code of Conduct does not specify student misconduct incident(s) associated with either Academic OSS or Tobacco OSS, so we omit these categories from the table. SDP incidents are based on the School District of Philadelphia's Student Code of Conduct for the 2012-13 school year. Information on PDE OSS categories retrieved from Pennsylvania Information Management System, User Manual (Volume 1), 2014-15.

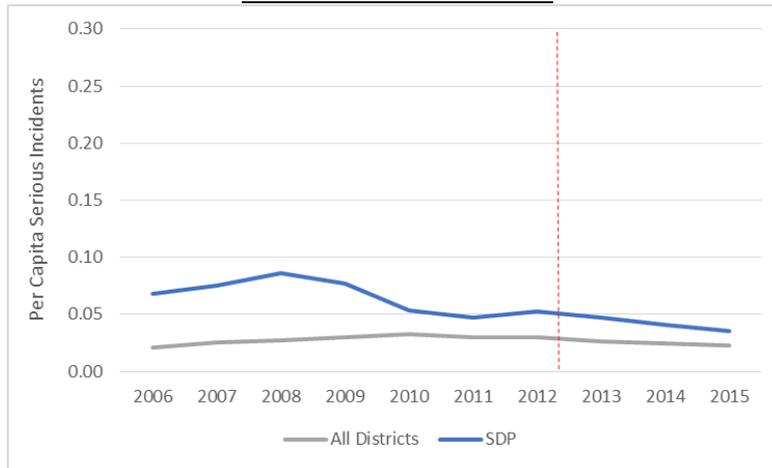
Figure A1. Distribution of Difference between Minimum and Maximum Total OSS



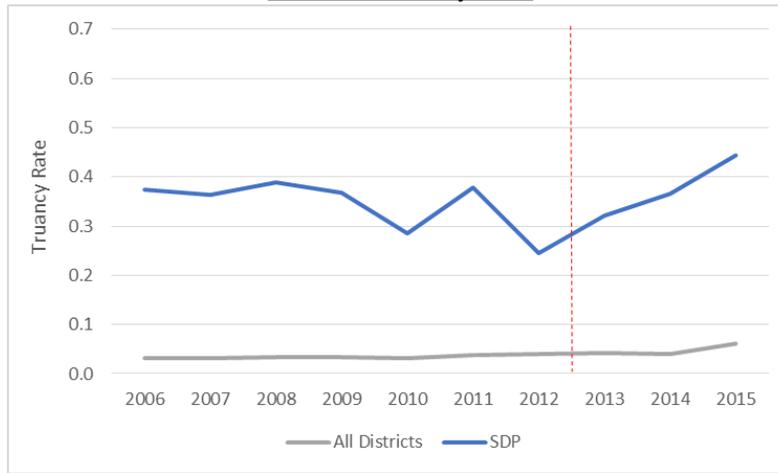
Note. Figure plots the difference between minimum per-capita OSS and maximum per-capita OSS for each district in the study period (i.e., 2005–2006 through 2014–2015 school years). Total OSS is inclusive of all OSS types (academic, conduct, drugs and alcohol, tobacco, violence, and weapons). One outlier district with a difference of 19 is not shown. For the 15 districts that are omitted from the analytic sample, the mean (sd) difference between the minimum per-capita OSS and maximum per-capita OSS over the study period is 2.01 (4.77), which represents a change of approximately 2 suspensions per student; for the 483 districts in our analytic sample, the mean (sd) difference between the minimum per-capita OSS and maximum per-capita OSS over the study period is 0.088 (.081), or approximately 9 suspensions per 100 students. Due to these extreme changes, and no indication of any relevant policy changes within the suspect districts during the time period, we determined that underlying reported OSS data for these districts were unreliable.

Figure A2. Trends in Serious Incidents, Truancy Rate, and Math and ELA Proficiency in Philadelphia and All Other Pennsylvania Districts

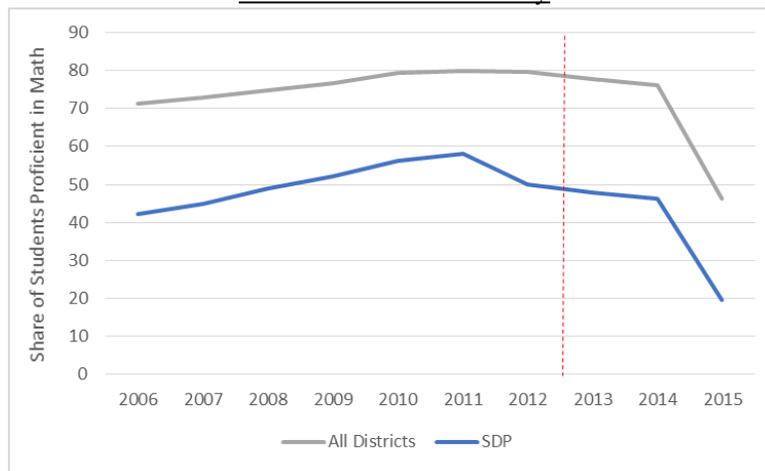
Panel A: Serious Incidents



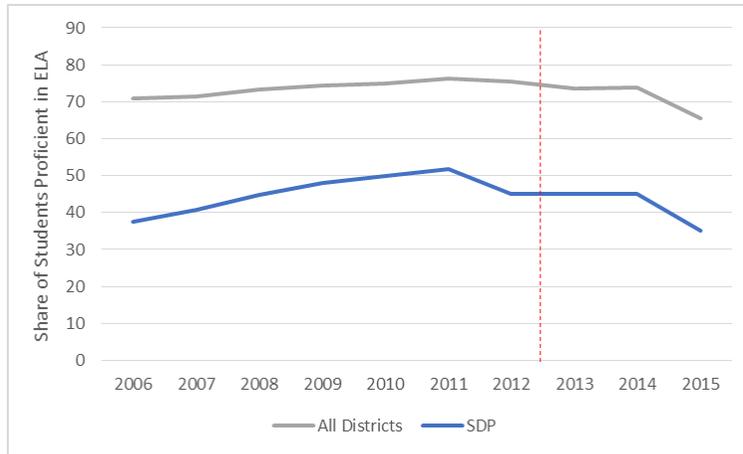
Panel B: Truancy Rate



Panel C: Math Proficiency



Panel D: ELA Proficiency



Note. The year 2006 refers to the 2005–2006 school year. SDP refers to the School District of Philadelphia; All Districts include all 482 comparison districts. The vertical line indicates the first year (2012–2013) of Philadelphia’s policy reform.