Disability as Discipline? Effects of the New York City Suspension Ban on Identification of Students with Disabilities*

Jo Al Khafaji-King†

New York University

January 2024

Abstract

This study evaluates the unintended consequences of the 2012 suspension ban in New York City. I find that the ban induced a substitution towards classification for students at high risk for suspension—Black students, male students, and those in schools with a high reliance on suspension. I find that disabilities that carry greater stigma and experience greater exclusion from the general education classroom drive the increases in classification. This substitution may benefit students if they are now receiving needed services. Simultaneously, ban-induced classifications may simply serve as a partial substitute for suspension.

JEL Codes: I24, I28
Keywords: school discipline reform, students with disabilities, policy evaluation

*This work was funded by the Institute for Education Sciences (IES) sponsored Predoctoral Interdisciplinary Research Training (PIRT) fellowship. The opinions expressed are those of the author and do not represent views of the IES, the U.S. Department of Education, the Research Alliance for New York City Schools, or the New York City Department of Education. I am grateful to Leanna Stiefel, Riley Acton, Austin Smith, Lucy Sorensen, Marina Gorzig, Tatiana Homonoff, Luis Rodriguez, Joseph Cimpian, and Kaitlyn O'Hagan for invaluable comments and support. I am grateful to participants, discussants, and chairs at the Association for Education Finance and Policy, Society for Research on Educational Effectiveness, and Association for Public Policy Analysis and Management annual conferences and members of the NYU Education Policy Work in Progress group and NYU IES-PIRT seminar for feedback on this work. Any remaining errors are my own.

†Contact: joalkhafajiking@gmail.com

1 Introduction

In recent years, widely-used exclusionary discipline practices in K-12 public schools (e.g., suspensions and expulsions) have been highlighted as detrimental to student outcomes (Welsh and Little, 2018a). Suspending or expelling a student causes lower graduation rates (Steinberg and Lacoe, 2018a; Sorensen et al., 2022), increased rates of incarceration (Deming et al., 2019), arrest (Mittleman, 2018), and contact with the juvenile justice system (Skiba et al., 2014; Sorensen et al., 2022), reduced academic achievement and increased absenteeism (Noltemeyer et al., 2015; Holt et al., 2022), and lower earnings (Davison et al., 2021). Moreover, research consistently documents that school personnel persistently and disproportionately discipline Black students, male students, students with disabilities (SWDs), and LGBTQ+ students (see Welsh and Little (2018b) for a comprehensive review of this literature). However, suspension disproportionality for SWDs has been contested (Morgan et al., 2019). Motivated by these negative effects on average and especially for historically marginalized groups, school districts and states have restricted or banned the use of suspensions for minor misbehaviors.

Recent evaluations of these bans generally estimate the direct effects of these reforms on suspension reduction and their indirect effects on student achievement for both high suspension risk students and their peers (Lacoe and Steinberg, 2018; Pope and Zuo, 2021; Craig and Martin, 2023). Other studies focus on whether the reforms were implemented with fidelity (Anderson, 2018; Steinberg and Lacoe, 2018b; Lacoe and Steinberg, 2018; Anderson et al., 2019), and how the reforms impacted disproportionality in discipline (Hashim et al., 2018; Baker-Smith, 2018; Wang, 2022; Cleveland, 2023), finding mixed results dependent upon the context. However, no research examines how high suspension risk students are managed in the wake of a suspension ban—what methods are school personnel using when suspension is no longer permitted? Understanding these potential alternative management strategies amidst an otherwise costless policy is essential for current and future suspension reforms to function more equitably and efficiently.

I investigate special education placement as a potential substitution for suspension in

the wake of New York City's 2012 reform that banned the use of suspension for minor offenses. Special education began as a way for students with disabilities (SWDs) to receive quality public education, but some argue that special education may also be used as a means to exclude Black students, contributing to the already entrenched racial inequities in education (Fish, 2019; Annamma et al., 2013; Blanchett, 2006). Moreover, the subjective nature of diagnosing many disabilities makes it difficult to disentangle behavior that is indicative of an underlying disability from that which is routine misbehavior. Indeed, "'it is essentially by definition that students with behavioral disorders engage in misbehavior at school more often than other students'" (Hurwitz et al., 2021).

I employ a dosage event study difference-in-differences strategy, relying on the fact that some schools and students were more treated by the ban than others due to a high prior reliance on suspension.[1] I define "more treated" groups in three ways: those students in above median suspension schools, Black students, and male students. Since these groups were subjected to higher suspension rates prior to the suspension ban, they are more likely to be subjected to alternative management strategies post-ban (i.e., special education classification) than those students in less treated groups. I also leverage variation across disability classification types based on their degree of exclusion and the stigma associated with the label to identify those classifications that would be a closest substitute to suspension. I ultimately follow Fish (2019)'s categorization of disabilities into low-status, stratified-status, and high-status classifications. Low-status classifications carry greater stigma relative to high-status classifications and students with low-status classifications are more likely to be excluded from the general education classroom than students with a high-status classification. Stratified-status classifications are more stigmatizing and exclusionary for Black students than non-Black students, hence the name "stratified." I confirm these patterns of exclusion in my own data and consider low-status classifications as the closest substitute for suspension given their higher levels of exclusion. I consider high-status classifications to be the least likely substitute for suspension.

I find that the 2012 suspension ban induced an increase in disability classifications,

[1] This strategy is also used in Ballis and Heath (2021a); Cleveland (2023); Craig and Martin (2023).

which was driven primarily by increases in low-status and stratified-status classifications. Increases in low-status classifications are driven primarily by students at high risk for suspension, with minimal change in disability classification rates for their respective comparison groups. Notably, while high-status classifications are increasing across my time period of analysis, there are no differential increases in high-status classifications for more treated groups relative to less treated groups. One exception is that Black students in low suspension schools are more likely to receive a high-status classification than their White peers, but Black students in high suspension schools are no more likely to receive a high-status classification than their White peers. Further, I find that male students in high suspension schools drive the increases in low-status and stratified-status classifications. Lastly, increases in low-status disabilities are driven by emotional disturbance classifications, with no change in intellectual disability classifications, which may be less malleable than emotional disturbance in that their diagnosis primarily relies upon academic performance and, historically, IQ tests (Boat et al., 2015).

While these increases in low- and stratified-status classifications may be an unintended consequence with unknown effects, evidence from prior studies suggests that the NYC ban caused improvements across a variety of dimensions and that the often-cited concerns with suspension bans in terms of worsened school climate and negative peer effects were not actualized (Craig and Martin, 2023). Moreover, if suspension previously prevented the marginal student from receiving needed special education services, this substitution effect may be beneficial for those newly classified students (Ballis and Heath, 2021b). With these findings in mind, my paper points to potential avenues through which an otherwise successful reform may be improved, as well as highlighting where special education may fall short or not function as intended. Further, my finding of increased racial disproportionality in SWD classification reiterates calls for culturally responsive reforms, which decenter color-blind responses, to address racial disproportionality in suspension (Welsh and Little, 2018b). Otherwise, disproportionalities in one arena (suspension) may reproduce in another (i.e., disability classification). Moreover, these results negate the assumption that suspension bans are costless policies, especially in the context of New York

City where special education provision is exceptionally costly. So long as special education may not be serving its intended purpose and instead functioning as a way to remove students, the ban that induced that substitution is not costless.

In addition to providing key evidence for education policymakers seeking to improve the implementation of discipline reforms and special education service provision, this paper draws on, and contributes to, three strands of related literature. First, I contribute to the literature regarding the efficacy of suspension restrictions. This includes evaluations of reforms in California (Wang, 2022), Massachusetts (Cleveland, 2023), New York City (Craig and Martin, 2023; Baker-Smith, 2018), Los Angeles (Pope and Zuo, 2021; Hashim et al., 2018), Arkansas (Anderson, 2018), Rhode Island (Craigie, 2022) and Philadelphia (Steinberg and Lacoe, 2018b). Each of these studies falls into three broad categories: implementation fidelity (Wang, 2022; Craigie, 2022; Anderson, 2018; Steinberg and Lacoe, 2018b), immediate effects on key student outcomes for both high-risk students and their peers (Cleveland, 2023; Craig and Martin, 2023; Pope and Zuo, 2021), and, to some extent, disciplinarians' substitution mechanisms (Wang, 2022). I build on the latter, which evaluates substitution to different suspension types as an alternative management strategy (Wang, 2022). However, I rely on a context (NYC) that had successful implementation without changes in other suspension types.[2]

Second, I contribute to literature regarding the theoretical and empirical links between suspension and disability classification. I draw from Disability Critical Race Theory—specifically tenet 1 regarding the interdependence of racism and ableism in upholding "notions of normalcy" (Annamma et al., 2013)—to frame my empirical results. I conceptualize special education classifications as tools of exclusion and "mechanisms of behavior conformity" (Cruz et al., 2021), specifically for Black students and students of color (Kim et al., 2010; Blanchett, 2010). While most of the prior literature on links between disability and suspension has been descriptive, I am, to my knowledge, the first to establish a causal link as to how disciplinarians similarly use suspension and special education: the absence of one

[2]Craig and Martin (2023) show that disciplinarians in New York City did not simply record lower-level infractions as higher-level infractions that would still allow for suspension, as was the case in California (Wang, 2022). I confirm this in my data in Figure 1.

(suspension) may induce substitution to the other (special education).

Lastly, my findings are relevant to the small, yet growing, literature on the causal effects of special education. The literature writ large suggests that special education, both broadly and for the marginal student, causes improvements across a variety of outcomes including increased educational attainment (Ballis and Heath, 2021b), improved academic outcomes (Schwartz et al., 2021), and reduced suspension likelihood (Hurwitz et al., 2021). Notably, however, Black students may be adversely affected by misclassification (Ballis and Heath, 2021a). It may indeed be the case that those students that receive a ban-induced classification benefit from additional services. Additionally, if it is true that Black students are under-represented in special education (Morgan et al., 2023, 2015), these classifications may be especially salient for Black students. Yet, if these ban-induced classifications are more similar to misclassification, we may expect adverse effects as observed in Ballis and Heath (2021a). Regardless, receiving a special education classification, irrespective of the status, may be an improvement over the counterfactual world in which the student was suspended. While I do not address these secondary outcomes in this paper, I point to directions for future work to derive a more normative judgment.

2 Background

2.1 Evidence on School Discipline Policy Reform

Zero-tolerance discipline in K-12 emerged in the 1970s and 80s as a byproduct of "tough-on-crime" legislation that employed broken-windows theory to justify harsh responses to minor misbehaviors in schools (Skiba and Losen, 2015). Proponents of zero-tolerance policies posit that punishing minor offenses such as truancy, profanity, or dress code violations would prevent more egregious offenses (e.g., fighting, aggression, etc.), suspension rates more than doubled in the years following the inception of tough-on-crime policy in schools, with a variety of negative consequences ensuing (see Welsh and Little (2018a) and Welsh and Little (2018b) for a comprehensive review of these consequences). While these effects were felt across all demographic groups, Black students, SWDs, and other

under-represented minorities bore the brunt of zero-tolerance policies, with disproportionality in suspension persisting today, even after the rollback of zero-tolerance practices under 2014 guidance by the Obama administration.

Modern-day disciplinary codes of conduct have begun to move away from the zero-tolerance paradigm, with schools, districts, and states implementing suspension bans and restrictions. These reforms are motivated by the need to counter the negative effects of suspensions on those suspended, and, in some contexts, the peers of those suspended (Steinberg and Lacoe, 2016). Opponents of reform cite the role of peer effects—disruptive students negatively impact the learning environment for their peers—as well as a need to ensure safety in the classroom (Carrell and Hoekstra, 2010; Carrell et al., 2018; Hwang and Domina, 2020).

The relatively nascent causal literature on suspension reform primarily documents their impacts on student outcomes. When reforms are implemented with fidelity, as in the case of New York City's ban on suspensions for minor offenses, test scores improve for all students due to improvements in school climate (Craig and Martin, 2023). Evidence from Massachusetts' efforts to reduce all suspension presents similar results, with the reform causing improvements in ELA test scores, absenteeism, and dropout rates (Cleveland, 2023). However, Baker-Smith (2018) documents that the NYC reform increased racial disproportionality in multiple suspensions for ninth grade Black girls, suggesting that the ban did not reduce disproportionality in suspension for some groups. Pope and Zuo (2021) provide evidence from Los Angeles, where reductions in suspensions caused reductions in test scores, on average—localized improvements for high suspension risk students were offset by small, yet widespread, impacts on low suspension risk peers.

Another strand of literature focuses on reform compliance. Prominent reforms in Philadelphia, Arkansas, and California experienced implementation challenges that limited the ability to evaluate the average impact of the bans. Anderson (2018) finds that Arkansas' ban on suspensions for truancy did not affect the schools the policy targeted—schools with high rates of truancy, suspension, and minority students were the least likely to comply. In fact, the reform, on average, was associated with increases in absenteeism, driven entirely

by schools that did not comply with the reform (Anderson et al., 2019). Evidence from Philadelphia tells a similar story of imperfect compliance (Steinberg and Lacoe, 2018b). For schools that successfully implemented the reform, there were slight improvements in achievement for students at high risk of suspension and a negligible impact on their peers. Peers in schools that only partially complied, however, experienced declines in achievement and attendance. Lastly, Wang (2022) presents evidence from restrictions on suspensions for willful defiance in a California school district. Suspensions for willful defiance were reduced, but suspensions for more severe infractions increased, even when the number of student infractions did not change. The substitution towards documenting infractions as more severe is interpreted as an unintended behavioral response from disciplinarians.

2.2 Special Education Services

The 1975 Education for All Handicapped Children Act, reauthorized as IDEA in 1990, provided students with disabilities with protections in schools, guaranteeing the right for students with disabilities to a free and appropriate public education (FAPE) as well as providing funding for states to implement special education. Subsequent reauthorizations of IDEA required that SWDs be educated in their least restrictive environment (LRE) and guaranteed services should they be evaluated to have a federally recognized disability (Katsiyannis et al., 2001). Additionally, the 1997 reauthorizations added protections regarding school discipline for SWDs. These provisions mandated that SWDs could not be placed in an alternative educational environment or suspended for more than 10 days without having a behavioral team determine whether the catalyzing behavior was due to a student's disability. However, for suspensions less than 10 days, SWDs were subject to the same regulations as their general education peers and a behavioral team assessment was not required (Katsiyannis and Maag, 1998).

IDEA also established 13 disability categories, which can be partitioned into "hard disabilities" (i.e., deafness, deaf-blindness, hearing impairment, orthopedic impairment, traumatic brain injury, visual impairment) and those that are more subjective. These sub-

jective disabilities are autism, emotional disturbance, intellectual disability, other health impairment, specific learning disability, and speech-language impairment (Fish, 2019). Due to the relatively subjective nature of these classifications, their designation is based on performance, student behavior, an adult's perception of that behavior, and the structures in place to address a student's behavior (Cruz et al., 2021). Dependent upon the context, some behavior may be perceived as a disability and met with services, or may be perceived as a student simply being "bad," resulting in suspension. Existing evidence documents the potential role that school context (e.g. student body racial composition) plays in the perception of a student's behavior, context that is outside of the control of the student and unrelated to the presence (or absence) of an underlying disability or behavioral issue (Okonofua and Eberhardt, 2015; Bal et al., 2019; Fish, 2019; Chin, 2021; Cruz et al., 2021; Elder et al., 2021; Stiefel et al., 2023).

Within subjective classifications, Fish (2019) identifies autism, other health, and speech-language impairment as high-status classifications; emotional disturbance and intellectual disability as low-status; and specific learning disability as stratified-status. Fish (2019) determines these groupings upon the stigma that a label carries, parental preferences for specific disability classifications, the services provided, and the degree of exclusion from the general education classroom. For example, high-status classifications are *not* stigmatized as indicating a lack of intelligence and, in some cases, carry a positive connotation. In the case of low-status classifications, emotional disturbance and intellectual disability are stigmatized as classifications for poorly behaved students or those with a "deficit in ... intelligence" (Fish, 2019). Students with low-status classifications spend less time with their general education peers and, as a result, may be unable to access the benefits that inclusive classrooms may provide (Myderrizi, 2023; Theobald et al., 2019). Stratified-status classifications are higher status for specific groups in that they have been used by advantaged families (i.e., White and high-income) to receive additional services, but function as a "dumping ground" for families of color (Blanchett (2010) via Fish (2019)).

While the intent of special education is to provide students with the services that they need to succeed, as exemplified by the tenets of IDEA, some argue that special education

has deviated from its intended purpose and instead functions as a "new legalized form of structural segregation and racism," as exemplified by Black over-representation in special education (Blanchett, 2006). Indeed, Chin (2021) documents the increase in Black students being referred to special education services in the wake of racial integration, and Fish (2019) documents the greater prevalence of Black students in special education in schools with a higher proportion of White students. Moreover, Black students are persistently over-represented in low-status classifications, whereas non-Black students are over-represented in high-status classifications (Blanchett, 2010).[3] In this vein, we can conceptualize special education classifications as potential mechanisms of exclusion, functioning to remove specific groups from the general education classroom. It is important to note, however, that conditional on teacher-rated behavior and social skills (Morgan et al., 2015), health status at birth (Elder et al., 2021), or kindergarten academic difficulties (Morgan et al., 2023), disproportionality is reversed such that Black students are under-represented in special education and are no more likely to be excluded from the general education classroom (Morgan et al., 2019).[4]

2.3 Policy Context

Under the Bloomberg administration, the New York City Department of Education (NYCDOE) updated the discipline code for the 2012-13 academic year to prohibit the use of suspension for disorderly behavior, also referred to as Level 2 infractions. Prior to the ban, students could be suspended for minor disruptions to the classroom, such as gambling, lying, smoking, leaving class without permission, or persistent non-compliant behavior. Non-compliant behavior includes having an unexcused absence from school, being tardy, or making too much noise in class or on school premises. Post-ban, it was recommended that this misbehavior be met with non-punitive interventions, such as restora-

[3] This is also observed in my data. See Table A2 for racial differences in disability classification.

[4] The merits of disproportionality estimates that control for these characteristics has been questioned (Collins et al., 2016; Skiba et al., 2016). Moreover, it may be the case that including specific covariates controls away variability that, in practice, may not be realistic. For example, teachers, because of implicit bias, systematically rate Black students' behavior and social skills different from White students. This means forcing a ceteris paribus comparison between Black and White students on this dimension creates an out-of-sample prediction that controls away bias that is relevant to disproportionality estimates.

tive practices, referrals to pupil personnel teams (counseling services), parent outreach, behavioral progress reports, Functional Behavioral Assessments (FBAs)/Behavioral Intervention Plans (BIPs), etc. While alternative practices were encouraged, the training that teachers received in implementing these practices was minimal and likely varied dependent on the culture of the school or an individual teacher's discipline philosophy. As such, these *recommended* practices were not necessarily *implemented* practices.

A concurrent evaluation by Craig and Martin (2023) shows that the ban was met with fidelity from schools — the suspension rate for Level 2 offenses was met with near perfect compliance and the overall rate of suspensions for other levels was also reduced, suggesting that the ban had spillover effects on reducing suspensions for higher level offenses. The authors also provide evidence that the ban improved reading and math test scores for both students that were and were not likely to be suspended prior to the ban, contradicting evidence on negative peer effects of suspension bans as found in Philadelphia (Steinberg and Lacoe, 2018b) and Los Angeles (Pope and Zuo, 2021). This improvement is attributed to improvements in school climate and, likely, the high compliance with the policy.

The NYCDOE context is also unique in that it serves a relatively high number of students with disabilities. Every student classified with a disability receives an individualized education plan (IEP), which is the result of a multi-step process from classification to provision of services. To receive an IEP, a student must be referred by either DOE school officials or a parent who believes that the student may have a disability. The referral is then sent to a child's school where parents are then asked to meet with the school social worker and IEP team to evaluate the child based on four manifestations: social history, psychoeducational metrics, classroom observation, and a physical examination. Independent assessments by non-NYCDOE personnel are also permitted, but the evaluation must be completed within 60 school days of the referral.

Upon completion of the evaluation, the IEP team will meet with the parent to construct an IEP to serve the needs of the child as demonstrated in the evaluation. The IEP itself contains nine components, including the students' current performance level, annual goals, progress reports, programs and services required, the level of inclusion, standardized test-

ing participation status, goal diploma (NYCDOE provides the option of alternative diplomas for SWDs), post-secondary transition services, and the language of instruction and service. Post-2012, all schools were required to provide accommodations for SWDs that meet their IEP requirements.[5]

2.4 Conceptual Framework

Broadly, the examination of increases in special education as a potential unintended consequence of suspension reform is motivated by the documented parallels between disability and exclusionary discipline in the extant literature. Some prior work shows that persistent misbehavior leads to a disability classification (Hurwitz et al., 2021), similar to how prior problem behavior predicts exclusionary discipline, on average (Huang, 2020). "Well-behaved" students with disabilities are more likely to be included with their general education peers, suggesting that misbehaved students with disabilities are more likely to be excluded (Anderson, 2021). While these studies do not document a causal relationship between discipline and disability, the evidence suggests that they are intricately related. As such, when examining the effect of a policy that affects suspensions, we should also expect the relationship between disability and suspension to change in tandem. Moreover, the persistent evidence regarding the disproportionate representation of SWDs in exclusionary discipline motivates examination of a potential causal relationship between discipline and disability (Welsh and Little, 2018b).

The NYC suspension ban fundamentally restructured how school staff could manage student behavior. When suspension is no longer viable, other alternatives must be used. Unless the ban itself drastically improved the behavior of students, teachers will still need a way to manage misbehavior in the classroom as there are incentives to remove disruptive students due to negative peer effects and to improve a teacher's classroom capacity (Lazear, 2001; Carrell and Hoekstra, 2010; Carrell et al., 2018; Steinberg and Lacoe, 2018a;

[5] I address this threat to validity through controlling for a school's average pre-ban SWD classification rate interacted with a linear time trend and conduct a robustness check to exclude students in most restrictive settings that would primarily be affected by the inclusion policy. Moreover, this policy affected *how* services were delivered to already classified SWDs, not necessarily *who* was classified.

Welsh and Little, 2018a; Pope and Zuo, 2021). Put simply, this means that post-ban, there is likely to be unmet demand for behavioral management strategies unless student behavior improved such that all disorderly behavior reduced to zero.

Many of the potential ways that teachers may meet this demand in the wake of a suspension ban include referrals to and interactions with behavioral teams. Indeed, the 2012 discipline code recommends the use of behavioral contracts/reports, referrals to pupil personnel teams, referral to counseling services, functional behavioral assessments (FBAs) and behavioral intervention plans (BIPs), and interventions by counseling staff. Craig and Martin (2023) also document that teachers began to remove students from just one class period (i.e., make office referrals). Each of these alternatives functions similarly to how a student may be classified with a disability. While teachers may not intend for their students to be designated with a disability, when required to substantially reform their management strategies, they may rely on behavioral teams to provide support in the classroom. As a byproduct, there is an unintended increase the number of students that interact with teams that have the ability to classify students with a disability. Assuming that these teams classify a consistent proportion of students as having a disability and are not able to directly observe the "true" underlying disability, this means that the total number of students with subjective disabilities should increase given the overall increase in contact between students and behavioral teams.

Within this increase in disability classification on average, there is likely heterogeneity in how increases are distributed between high, stratified, and low-status classifications. On one hand, we may expect that suspension made it difficult for students to receive services that they needed—once students receive services, they become less likely to be suspended relative to not receiving services (Hurwitz et al., 2021), even if students with disabilities, on average, are more likely to be suspended than their general education peers. Prior to the ban, teachers may have relied too heavily on suspension, diverting students away from receiving services and, instead, labeling them as "bad", rupturing the connection that a student had with a school, and inducing a self-fulfilling prophecy of misbehavior (Kennedy-Lewis and Murphy, 2016). In the absence of suspensions, however, students are

diverted towards teams that can determine and, ultimately do, provide needed services. Indeed, Barnard-Brak et al. (2023) demonstrates that approximately 1 in 14 expelled students have an unserved disability. If this hypothesis holds true, we should expect that classifications, especially those with greater inclusion and less stigma, should increase as it is less likely that these services are functioning to remove specific students from the classroom—students who do not carry a high-stigma label are less likely to be excluded from the classroom.

There is also the potential for disability classification, specifically those that carry high stigma and serve students in more restrictive environments, to function as a method of exclusion, i.e., a partial substitute for suspension. Certain disability classifications are often used as mechanisms to exclude students, specifically Black students (Blanchett, 2010; Fish, 2019), and both suspension and subjective disability classifications are influenced by student misbehavior (Bal et al., 2019; Cruz et al., 2021; Hurwitz et al., 2021). As such, we may expect specific subgroups to be disproportionately classified with high stigma, less inclusive classifications, classifications that may have deviated from the intended purpose of special education and now serve as a "dumping ground" (Blanchett, 2010) for students whose behavior would otherwise be met with suspension.

3 Data

3.1 Administrative Data

I use student-level, administrative data provided by NYCDOE ranging from the 2007/08 to 2014/15 school year. I limit my estimation to pre-2015 years due to a second wave of discipline reform that began in the 2015/16 school year. The data include information on a student's grade, school attended, standardized middle school math and reading test scores,[6] demographic characteristics (race/ethnicity, sex, FRPL status, ELL status), whether a student was classified as a SWD, and their specific classification, if applicable.

[6] I use a time invariant measure of reading and math achievement, equivalent to the mean of a student's pre-ban middle school test scores. Implicitly, this requires that students must have at least one test score to be included in the analytic sample.

I include all students in middle and high schools in New York City, but exclude charter schools from the sample.[7] Ultimately, the sample covers over 500,000 students in over 700 middle and high schools throughout the city.

NYCDOE did not collect information on disability type in the 2012-13 school year, but they did collect information on whether a student had any disability classification. To determine the disability type in 2013, I first assume a consistent disability classification for students that have the same classification in 2012 and 2014. After this process, I use a multinomial logit utilizing all non-charter middle and high-school students, with student and school characteristics, to predict the classification (either high, low, stratified, or hard) that is most likely based on the student's 2014 and 2015 data. This process applies to less than 1 percent of all observations, 97.2 percent of which are sorted into a stratified-status classification, with the remaining 2.8 percent allocated to high-status classifications.[8]

3.2 Descriptive Statistics

Descriptive statistics and 2011/12 baseline differences in disability classification rates are presented in Tables 1 and A1, respectively. Hispanic students are the largest racial group in New York City, making up 40 percent of the student body, followed by Black students (30 percent), Asian students (16 percent), and White students (13 percent). Nearly 16 percent of students in NYCDOE are classified with a disability. 9.8 percent are classified with a stratified-status disability; 4 percent, high-status; 1.2 percent low-status; 0.03 percent hard classifications. Male students and students in high suspension schools are more likely to have any disability classification, but Black students are more likely to have stratified and low-status classifications and less likely to have high-status classifications.

Table A2 breaks down the specific classification type for each status category. Of high-

[7] Charter schools are subject to their own disciplinary regime and are, therefore, not subjected to the suspension ban. Moreover, they were not required to accept SWDs prior to the policy change, making them an implausible comparison group. Additionally, I only use schools that reported suspension rates in the year before the ban, limiting the influence that newly opened schools may have on results, while also allowing for inclusion of schools that opened prior to the ban but were not observed in all years. This restriction is also key, as I use the 2011-12 school-level suspension rates to define treatment. Ideally, student-level suspension data would be used. See Section 4 for further discussion of this limitation.

[8] These imputations do not drive my results. I conduct a robustness check in FigureA3 that uses the non-imputed SWD indicator as my outcome, finding that my results for all specifications align.

status classifications, other health impairment (primarily ADHD) is the most common, and of low-status classifications, emotional disturbance is most common. Specific learning disability is the most common classification of all. Students with low-status classifications, on average, are disproportionately served in self-contained classrooms—low-status classifications comprise only 7.77% of classifications, but 13.79% of classifications served in self-contained settings. Those with stratified-status classifications are less likely to be excluded on average. High-status classifications are generally proportionately excluded. SWDs in above-median schools, Black SWDs, and male SWDs are all more likely to have low-status disabilities. For Black and male students, this risk is nearly double that of their respective comparison groups. Conversely, Black students and students in high-suspension schools are generally less likely to have high-status classifications.

While male SWDs are less likely to have a high-status classification, of those that do receive high-status classifications, male SWDs are more likely to be in self-contained classrooms than female SWDs with high-status classifications. Black SWDs are more likely, regardless of status, to be in self-contained classrooms. SWDs in above-median suspension schools are also more likely to be excluded regardless of classification, with the exception of those with an intellectual disability classification. Overall, these descriptive facts largely align with the criteria presented by Fish (2019). As such, I use these groupings to produce my primarily results and conduct robustness checks using these individual classification types.

4 Empirical Strategy

4.1 Suggestive Evidence: Examining Trends across Time

I exploit the sharp timing of the suspension ban to provide plausibly exogenous variation in the number of students suspended in a year, which I hypothesize induces a substitution towards identification of students with disabilities.[9] To provide suggestive evidence

[9]Figure 1 provides visual confirmation of this drop by infraction level (Panel A) and on average for above and below median suspension schools (Panel B).

regarding this hypothesis, I start by documenting the year-over-year change in SWD classification rates using the following model:

$$Y_{\text{isgt}} = \beta + \alpha_{\text{sg}} + \sum_{k \neq 2012}^{2015} \rho_k 1(t = k) + X_{\text{isgt}}\phi + X_{\text{st}}\eta + \epsilon_{\text{isgt}}$$

where Y_{isgt} is a dichotomous outcome variable that indicates whether a student i in school s and grade g at time t was classified with a disability (or a specific disability classification type).[10] I include school by grade fixed effects (α_{sg}) such that the estimates reflect changes in SWD classification within a school-grade instead of on average across the entire NYC school district. As such, these estimates are akin to estimating the change in a school-grade's classification rate when suspension rates are lower than average (i.e., post-ban). Student-level controls, including sex, race, ELL status, FRPL status, years present in the administrative data, and average pre-ban reading and math scores are also included.[11] I also include a linear time trend interacted with school-level average shares of students with disabilities and, in some models, the school-level average share of students suspended in the pre-ban to control for underlying trends in these variables. The coefficients of interest are ρ_k, which represent estimates of the change in SWD classification relative to 2012 within a school-grade.

4.2 Causal Estimates: Event Study Difference-in-Differences

To identify the causal effect of the ban, I exploit the fact that some students were affected more by the suspension ban than others (i.e., some students received a higher treatment "dosage"), following a strategy used in Craig and Martin (2023) and Cleveland (2023) and, to an extent, in Ballis and Heath (2021a). First, I replicate Figure 1, Panel B the same drop in suspensions that Craig and Martin (2023) observe. Upon confirming my source of variation, I estimate "dosage" difference-in-difference models where treatment status

[10] The primary estimates use the low-, stratified-, and high-status groupings, but I also provide estimates for individual disability classifications in Appendix Figure A1.

[11] I also provide correlation coefficient (Table A5 between mean test scores and a consistent measure for all students, which is the last test score available prior to the policy change.

is determined by whether a school is above the 2011-12 district median suspension rate of 6 percent. Schools that relied more heavily on suspension than average would be forced to reduce their suspension use more so than a school that barely suspended students even when suspension was an option. Therefore, these schools would be more likely to increase their SWD classifications as a response to the ban. Ideally, I would be able to assign treatment status given a school-grade's suspension rate for Level 2 infractions specifically as in Craig and Martin (2023), however the data available do not allow for simultaneous examination of specific disability classification and individual suspension records.[12] As such, I rely on suspension data aggregated to the school-level to determine treatment status. This means that there is substantial noise in the assignment to treatment. For example, a school could be identified as treated, but their suspensions consist of only Level 3 infractions. This school would be relatively unaffected by the ban, but would be mis-identified as treated due to their high reliance on suspension on average. Therefore, the estimates presented are attenuated and are likely a lower bound for effect sizes. In addition to some schools being more subject to treatment than others, certain demographic groups are also more treated.[13] Specifically, Black students and male students are disproportionately represented in suspensions. As such, I compare effects for Black students relative to their non-Black peers and effects for male students relative to their female peers.[14] I estimate variants of the following equation:

$$Y_{\text{isgt}} = \beta + \alpha_{\text{sg}} + \sum_{k \neq 2012}^{2015} \rho_k [1(t = k) \times 1(\text{High Risk}_{\text{isgt}})] + X_{\text{isgt}}\phi + X_{\text{st}}\eta + \epsilon_{\text{isgt}}$$

where ρ_k are still the coefficients of interest, where "High Risk" indicates either being in

[12] I use two mutually exclusive data sources in this paper. I use data from the Research Alliance for NYC Schools (RANCYS) in Figure 1, Panel A, which contains individual-level information on suspensions and infractions. However, RANCYS does not have reliable pre-2012 information on SWD classification type. Data provided directly by NYCDOE has reliable information on SWD classification type but they have not granted access to individual suspension data.

[13] This is not a traditional event study difference-in-differences estimation strategy in which there is a "clean" comparison group that is completely untreated. Nevertheless, the use of the event study difference-in-differences framework is helpful to understand how the ban differentially affected students in contexts that were more affected relative to those that were less affected (Craig and Martin, 2023; Cleveland, 2023).

[14] Hispanic students are not over-represented in suspensions and have similar suspension rates to White students. To better represent the New York City context, I include White, Asian, Hispanic, and Other students in my comparison group when investigating Black disproportionality in discipline.

a high-suspension school, or being Black, or being male.[15] These coefficients can be interpreted as the difference-in-difference estimate—that is, the difference in the classification rates between groups, relative to the 2012 difference between groups. These estimates can be interpreted as causal if we believe that SWD identification continued in parallel between treated and untreated groups, and that any deviation is driven by the ban and no other contemporaneous policy changes. While these assumptions are untestable, I do provide suggestive evidence that they are satisfied as evidenced by insignificant pre-trends for most years leading up to the policy change.

One such policy change that poses a threat to identification is the inception of a Shared Pathway to Success (aSPtS), which began in the 2012-13 academic year. Among other goals, this policy aimed for SWDs to be educated with their general education peers, using the general education curriculum to the greatest extent possible, and required that a SWD's zoned school serve their needs. These new requirements primarily affected students in more restrictive environments (i.e., "District 75", special education only schools) and already classified SWDs, which, in my primary models, do not contribute to new identifications. Broadly, aSPtS did not affect who was classified, instead, it affected how services were delivered. I conduct a robustness check to account for this policy change in Section V.C by removing students who were primarily affected by aSPtS—those in most restrictive settings (special education only schools)— finding that the main results are largely unaffected.

5 Results

5.1 Trends Across Time

Figure 2, Panel A and Table 2 contain estimates relative to the baseline rate of SWD classification in the 2011-12 academic year. On average, the proportion of students classified with subjective disabilities increased post-ban. Stratified-status and high-status classifi-

[15] I also estimate a specification where treated is a continuous variable equal to a school's 2012 suspension rate standardized to a mean 0 and standard deviation of 1, similar to the dose-response design used in Ballis and Heath (2021a). Figure A2 depicts these results.

cations drive these increases, with stratified-status classifications increasing by 1.7 percentage points (pps) from the baseline 2012 classification rate of 9.4 percent, and high-status classifications increasing by 1pp from the baseline rate of 4.1 percent. While small in absolute magnitude, from their baseline, high- and stratified-status classification rates increased by 24 and 18 percent, respectively. Low-status classifications also increased on average, increasing by 0.4pp from the baseline classification rate of 1.2 percent, or a 33 percent increase. I do not interpret these estimates as causal, since this interpretation relies upon the assumption that conditional on controls, the change in classification rates post-2012 is due solely to the policy change. Put differently, absent the ban, the trend in identification of SWDs would continue on the same path. With these trend estimates, there is evidence that prior to the suspension ban that classification rates were declining. Given this trend, it may be the case that the estimates for the effect of the ban are attenuated or represent a reversion to the mean.

5.2 Effect on SWD Identification

Table 3, Panel A and Figure 2, Panel B present estimates for the differential effect of the suspension ban on schools with high and low suspension reliance. Relative to low-suspension schools in 2012, students in high-suspension schools in 2015 were also more likely than students in low-suspension schools to be classified with a subjective disability. Differential increases in classification rates are driven by stratified-status and low-status classifications, classifications that possess greater stigma and exclusion from the general education classroom. The gap in classification rates between low- and high-suspension schools increases by 0.6pp for stratified-status classifications by 2015 (19 percent from the baseline difference of 3.1 percent) and 0.5pp for low-status classifications in 2015 (71 percent from the baseline difference of 0.7 percent), relative to the 2012 baseline difference. Increases in low-status classifications are driven primarily by high-suspension schools—there is little increase in low-status classifications for low-suspension schools. No estimates are significant for high-status classifications — low- and high-suspension schools increase their high-status classifications at similar rates relative to 2012. These estimates suggest

that the schools in which we would expect a substitution effect to occur are, indeed, potentially induced to using low- and stratified-status classifications to manage student behavior.

5.3 Treatment by Race and Sex

Next, I estimate specifications using being Black or male as a treatment since Black and male students are more affected by the suspension ban than their peers. This is because Black and male students are subjected to higher rates of discipline than their observationally similar peers.[16] In addition to these estimates being understood as difference-in-difference estimates, we can also think of these estimates as racial and gender disproportionalities in classification rates.

Table 3, Panel B and Figure 2, Panel C present estimates comparing Black students to non-Black students. At the 2012 baseline, Black students are more likely to be classified with stratified and low-status disabilities and less likely than non-Black students to be classified with a high-status disability. By 2015, the gap in disability classification between Black and non-Black students widened by 0.7pp for stratified-status classifications and by 0.5pp for low-status classifications, approximately 30 and 45 percent from their respective baseline differences. There is no differential change in high-status classifications between Black and non-Black students. In line with prior literature, Black students are more likely to be labeled with high stigma, more restrictive classifications especially when other exclusionary methods are not available, namely, suspension. Lastly, despite Black students' over-representation in suspension, they remain just as likely as their non-Black peers to experience an increase in the probability of a high-status classification.

I also provide estimates using gender as treatment in Table 3, Panel C and Figure 2, Panel D. Indeed, in 2012, male students were more likely to receive any subjective classification type than their female peers. Subsequent to the ban, this gender disproportionality in subjective disability classification increased, driven, perhaps unsurprisingly, by stratified-status and low-status classifications. Like Black students, male students are no

[16] This is true throughout the literature, as well as in the NYC context as demonstrated by Craig and Martin (2023)

more likely to receive a high-status classification after the ban than their female peers, except in 2013. I interpret these results with caution given significant pretrends.

5.4 Triple Difference Estimates

Lastly, I estimate two triple difference specifications: first, one that interacts year indicators with being Black in a high suspension school and, second, one that interacts year indicators with being male in a high suspension school.[17] While there is greater gender balance across above and below median suspension schools, there is a concentration of Black students in above median schools.[18] Estimates for the triple difference by school type and race are in Table 4, Panel B and Figure 3.[19] These estimates provide suggestive evidence that the result that Black students are more likely to be in stratified and low-status classifications is driven by high-suspension schools, although these differences are not statistically significant at conventional levels. Additionally, the results suggest that Black students in low-suspension schools are increasingly more likely than White students to be classified with a high-status disability, whereas Black and White students experience an equal increase in likelihood of high-status classifications in a high-suspension school. For schools with less punitive environments, it appears as though students that are over-represented in suspension begin to receive services. However, over-represented students are simultaneously more likely to receive low-status classifications. Overall, these results show that the hypotheses that suspensions were prohibitive to service reception and that some classifications are also excluding students are not mutually exclusive.

To further investigate whether the effects observed are due to special education being a

[17] I estimate both a fully interacted triple-difference specification as well as specifications for Black-White/Asian gaps and male/female gaps separately for high and low suspension schools.

[18] Estimates that use non-Black students as a comparison group are difficult to interpret due to extreme sorting based on race/ethnicity in high- and low-suspension schools. High suspension schools are disproportionately Black and Hispanic, whereas low-suspension schools are disproportionately White and Asian. As such, the non-Black comparison group in high-suspension schools consists of primarily Hispanic students, whereas the comparison group in low-suspension schools is predominantly White. The sorting of Black students into high-suspension schools, as well as the changing of the comparison group composition makes these estimates difficult to interpret, even with school-grade fixed effects. Instead, I use White and Asian students as the comparison group to maintain greater consistency of the composition of the comparison groups in high vs. low suspension schools, as well as allow some comparability with estimates that are common in the literature (i.e., Black-White disproportionality).

[19] Appendix Table A3 present these results between Black and non-Black students.

substitute for suspension, I estimate the ban's effect on males in above median suspension schools.[20] Figure 4, Panel B and Table 4, Panel A present these results. The difference in classification rates between males and females grows more in high-suspension schools than low-suspension schools. The same pattern occurs for stratified-status classifications, although it is not statistically significant. Notably, there is no difference in high-status classification rates between males and females in high-suspension schools and males and females and low-suspension schools. Taken together, there is suggestive evidence that the effects are most prominent where we would expect (high-suspension schools) and, within these contexts, for students that are more affected by the suspension ban.

5.5 Robustness

For special education classification to be a plausible substitute for suspension, there must be an overlap in how both disabilities and suspensions are assigned. In short, the observed increase in special education classifications should only be due to increases in classifications that are subjectively determined and there should be no change in the prevalence of hard classifications, which are more objectively determined. As a placebo test, I estimate the effect of the suspension ban on the incidence of hard classifications, which includes deafness, blindness, deaf-blindness, orthopedic impairment, traumatic brain injury, and hearing impairment.

I re-estimate all models: trends across time, high- and low-suspension treatment, Black vs. non-Black students, and male vs. female students. Table 5 presents these estimates. For all models, there is no change in hard classifications post-ban, with point estimates and 95 percent confidence intervals being very close to zero. This suggests that the persistent result of increases in disability classifications post-ban are not spurious correlations and are, indeed, due to changes in how behavior is managed in a school. Additionally, when estimating the impact of the ban by specific disability type (Figure A1), I find that emotional disturbance drives the estimates for low-status classifications, with little change

[20] I estimate models separately for above and below median schools (Figure 4, panels A-C) and also estimate a fully interacted triple-differences specification (i.e., interacting year, an indicator for being male, and an indicator for being in a high suspension school), which is depicted in panel D and Table 4, Panel A.

in intellectual disability. Emotional disturbance is exceptionally malleable, with criterion being as simple as "an inability to maintain satisfactory relationships with teachers and peers" (USDOE, 2018). Conversely, the criteria for intellectual disability is "significantly subaverage general intellectual functioning, existing concurrently with deficits in adaptive behavior and manifested during the developmental period" (USDOE, 2018). As such, we may expect the more malleable category to be more affected by the ban, which I observe in Figure A1, Panels B, D, and F.

I also estimate an alternative specification of treatment intensity using a continuous measure of suspension rates instead of a dichotomous measure. I use a school's 2011-12 suspension rate standardized to a mean of 0 with a standard deviation of 1 and find that low and stratified-status classifications increased more in schools with higher suspension rates (Figure A2). Therefore, the results presented throughout this paper are not simply a byproduct of a chosen treatment specification. Additionally, to alleviate concerns regarding imputation of classification type in 2013, I estimate all difference-in-difference models using an indicator for whether a student has *any* disability, which was collected in all years by NYCDOE. I find the same patterns, with classifications increasing more so for treated students than comparison students (Figure A3).

A potential threat to validity is the introduction of aSPtS, which was a program that was also implemented in 2012 with the intention of increasing inclusion of SWDs in the general education classroom, improving the academic achievement of SWDs, and increasing schools' capacity to serve SWDs. The reform primarily targeted SWDs in the most restrictive environments (i.e., District 75 schools, which are special education only) and affected how services were delivered to SWDs, not who was classified. Evidence suggests that the initiative improved inclusion in elementary and middle schools, but not in high schools (Stiefel et al., 2021). Put simply, the reform did not affect *who* received services, rather, it affected *how* students received services.

To ensure that the results presented are not due to aSPtS, I exclude students that ever attended a District 75 school prior to the ban. This excludes approximately 32,000 observations. Estimates using this restricted sample produce nearly identical estimates as the

main results, albeit slightly attenuated, as expected. The one exception is for male students relative to female students—the finding of an increased likelihood of a low-status classification does not remain robust. However, estimates for stratified-status classifications remain robust, as well as all estimates for Black students relative to non-Black students and for high suspension schools relative to low suspension schools. All estimates are presented in Appendix Table A3.

6 Limitations

This work is not without limitations. First, I am unable to derive a clean comparison group such that the estimates presented in this paper may be attenuated. Charter schools create their own disciplinary codes and were thus not subjected to the ban; however, these schools were also not required to accept SWDs prior to 2012. As such, these schools are fundamentally different from NYCDOE schools. Estimates generated from using charters as a comparison group would also be exceptionally attenuated since charter schools experience an influx of SWDs once required to accept these students. This occurs at the same time as the suspension ban.

Second, the data that I rely on does not contain suspension records for each student, but only contains school-level data. Data that contain suspension records for each student does not contain reliable information on students' disability type prior to 2012. This means that I am not able to define a granular level of treatment as in Craig and Martin (2023), but must rely on a broader treatment level. If anything, however, this broader definition of treatment likely means that estimates are attenuated—students in high suspension schools, but themselves are not at high risk for suspension, are defined as treated but unlikely to be affected by the ban. I attempt to examine this within school variability in the dosage of treatment through the triple difference that examines how disability classification rates for Black students and male students, i.e., high risk demographic groups, in high suspension schools change due to the ban.

Third, while I can address the potential issue of aSPtS increasing classifications, es-

timating whether the ban induced students into more restrictive environments is confounded by the inclusion initiative. This is due to the fact that more restrictive environments were reduced following 2012. Therefore, any specification that examines the impact of the ban on inclusion would be misattributed.

Lastly, this paper is unable to make a normative judgement as to whether this inducement of SWD classification is beneficial or harmful to students. I leave this question to further research.

7 Discussion

Understanding how discipline bans function, and their potential pitfalls, is of utmost importance to ensure these reforms accomplish their intended purpose of reducing suspension, improving school culture, and, at best, improving the achievement of not only students at high risk of suspension, but also their peers. The results presented in this paper build upon the literature that suggests that suspension bans, when implemented with fidelity, improve outcomes (Craig and Martin, 2023; Cleveland, 2023; Steinberg and Lacoe, 2018a). Specifically, I show that even when suspension bans achieve their primary goal of reducing suspension, that there are still methods that disciplinarians may be using to exclude students from the general education classroom, namely, disability classifications. Additionally, it may be the case that suspensions kept students from receiving services that were necessary for their success.

All students experienced increases in stratified-status classifications, yet Black students and male students were more likely than their non-Black or female peers to receive these classifications. This finding is in line with prior literature that suggests that stratified-status classifications are a "dumping ground" for children of color. Moreover, it may be the case that Black students that receive this classification are less likely to receive the same services as their non-Black peers, as noted by Blanchett (2010) and Fish (2019). Conversely, the documented increase in high-status classifications for Black students in low-suspension schools provides suggestive evidence that suspension may have prevented stu-

dents from receiving services. However, the disproportionate increase in stratified-status classifications concentrated on students at high risk for suspension provides evidence that, for some students, these classifications are used to potentially exclude them from the classroom or maintain the label of "misbehaved."

Perhaps more concerning is the disproportionate increase in low-status classifications for Black students and male students in high-suspension schools. These classifications possess high stigma and these results suggest that the label of "low-status" may simply replace the label that being suspended carried. The increased disproportionality in these classifications are driven almost entirely by Black students or students in high suspension schools — there is minimal increase in low-status classifications for their comparison groups. The hyper-concentrated increase in low-status classification on groups that are historically over-represented in suspension is the most suggestive evidence of low-status classifications being a partial substitute for suspension.

The results as a whole are consistent with prior literature that documents an association between disability classification and suspension. I provide a causal link between these two behavioral outcomes, suggesting that the absence of one management strategy (suspension) results in a substitution to another (special education). However, I do not take these results as evidence that the suspension ban did not work. While the consequence of increased classification may have been unintended, there is evidence that some of these services were needed, in line with Barnard-Brak et al. (2023).

With suspension, the intention is to exclude students, at its best, in order to improve the outcomes of peers that are experiencing disruption and, at its worst, as a manifestation of racial bias. The purpose of exclusion is apparent with suspension and the intention is often congruent with the outcome. With special education, however, the intention began as a way to provide students with the resources that they needed to learn—students with disabilities would be provided with equal educational opportunity. The intention is the opposite of that of exclusion, as evidenced by the LRE requirement of IDEA, yet there is evidence that, in some cases, special education has deviated to serve as a method of exclusion, functioning more similarly to suspension for some students.

Future policy should explicitly consider the potential avenues that disciplinarians may take to manage student behavior when a relied-upon strategy is no longer permitted. New York City's ban is one of the few that achieved its intended purpose of reducing suspension rates with minimal consequences in terms of worsened academic achievement of peers or worsened school climate. However, even the most efficacious reforms still present costly unintended consequences when general equilibrium effects are not considered — if student behavior does not change drastically, there will still be demand for suspensions that cannot be met when disciplinarians are not explicitly trained in alternative techniques. This paper provides evidence that a partial substitute may be special education classification. In some cases, this substitution may be beneficial, in which the provision of needed services through IEPs may be a potential tool to consider in future reforms. In other cases, however, this substitution may simply behave as a method to exclude students, albeit in a different manner than suspension. Ultimately, policy in one arena must consider potential spillover effects into others if equity goals are to be realized.

References

Anderson, K. P. (2018). Inequitable compliance: Implementation failure of a statewide student discipline reform. *Peabody Journal of Education 93*(2), 244–263.

Anderson, K. P. (2021). The relationship between inclusion, absenteeism, and disciplinary outcomes for students with disabilities. *Educational Evaluation and Policy Analysis 43*(1), 32–59.

Anderson, K. P., A. J. Egalite, and J. N. Mills (2019). Discipline reform: The impact of a statewide ban on suspensions for truancy. *Journal of Education for Students Placed at Risk (JESPAR) 24*(1), 68–91.

Annamma, S. A., D. Connor, and B. Ferri (2013). Dis/ability critical race studies (discrit): Theorizing at the intersections of race and dis/ability. *Race Ethnicity and Education 16*(1), 1–31.

Baker-Smith, E. C. (2018). Suspensions suspended: Do changes to high school suspension policies change suspension rates? *Peabody Journal of Education 93*(2), 190–206.

Bal, A., J. Betters-Bubon, and R. E. Fish (2019). A multilevel analysis of statewide disproportionality in exclusionary discipline and the identification of emotional disturbance. *Education and Urban Society 51*(2), 247–268.

Ballis, B. and K. Heath (2021a). Direct and spillover effects of limiting minority student access to special education.

Ballis, B. and K. Heath (2021b). The long-run impacts of special education. *American Economic Journal: Economic Policy 13*(4), 72–111.

Barnard-Brak, L., T. Stevens, and A. Kearley (2023). Expelled students in need of special education services using bayes' theorem: Implications for the social maladjustment clause? *Behavioral Disorders*.

Blanchett, W. J. (2006). Disproportionate representation of african american students in special education: Acknowledging the role of white privilege and racism. *Educational Researcher 35*(6), 24–28.

Blanchett, W. J. (2010). "telling it like it is: The role of race, class, & culture in the perpetuation of learning disability as a privileged category for the white middle class.". *Disability Studies Quarterly 30*(2), Article 2.

Boat, T. F., J. T. Wu, E. National Academies of Sciences, Medicine, et al. (2015). Clinical characteristics of intellectual disabilities. In *Mental disorders and disabilities among low-income children*. National Academies Press (US).

Carrell, S. E., M. Hoekstra, and E. Kuka (2018). The long-run effects of disruptive peers. *American Economic Review 108*(11), 3377–3415.

Carrell, S. E. and M. L. Hoekstra (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone's kids. *American Economic Journal: Applied Economics 2*(1), 211–228.

Chin, M. J. (2021). Desegregated but still separated? the impact of school integration on student suspensions and special education classification.

Cleveland, C. (2023). Rethinking discipline: The effects of state discipline reform laws on students.

Collins, K. M., D. Connor, B. Ferri, D. Gallagher, and J. F. Samson (2016). Dangerous assumptions and unspoken limitations: A disability studies in education response. *Multiple Voices for Ethnically Diverse Exceptional Learners 16*(1), 4–16.

Craig, A. and D. Martin (2023). Discipline reform, school culture, and student achievement.

Craigie, T.-A. (2022). Do school suspension reforms work? evidence from rhode island. *Educational Evaluation and Policy Analysis 44*(4), 667–688.

Cruz, R. A., S. S. Kulkarni, and A. R. Firestone (2021). A quantcrit analysis of context, discipline, special education, and disproportionality. *AERA Open 7*.

Davison, M., A. M. Penner, E. K. Penner, N. Pharris-Ciurej, S. R. Porter, E. K. Rose, Y. Shem-Tov, and P. Yoo (2021). School discipline and racial disparities in early adulthood. *Educational Researcher 51*(3), 231–234.

Deming, D., A. Bacher-Hicks, and S. Billings (2019). The school to prison pipeline: Long-run impacts of school suspensions on adult crime.

Elder, T. E., D. N. Figlio, S. A. Imberman, and C. L. Persico (2021). School segregation and racial gaps in special education identification. *Journal of Labor Economics 39*(S1), S151–S197.

Fish, R. E. (2019). Standing out and sorting in: Exploring the role of racial composition in racial disparities in special education. *American Educational Research Journal 56*(6), 2573–2608.

Hashim, A. K., K. O. Strunk, and T. K. Dhaliwal (2018). Justice for all? suspension bans and restorative justice programs in the los angeles unified school district. *Peabody Journal of Education 93*(2), 174–189.

Holt, S. B., K. Vinopal, H. Choi, and L. C. Sorensen (2022). Strictly speaking: Examining teacher use of punishments and student outcomes.

Huang, F. L. (2020). Prior problem behaviors do not account for the racial suspension gap. *Educational Researcher 49*(7), 493–502.

Hurwitz, S., E. D. Cohen, and B. L. Perry (2021). Special education is associated with reduced odds of school discipline among students with disabilities. *Educational Researcher 50*(2), 86–96.

Hwang, N. and T. Domina (2020). Peer disruption and learning: Links between suspensions and the educational achievement of non-suspended students. *Education Finance and Policy*, 1–44.

Katsiyannis, A. and J. W. Maag (1998). Disciplining students with disabilities: Issues and considerations for implementing idea '97. *Behavioral Disorders 23*(4), 276–289.

Katsiyannis, A., M. L. Yell, and R. Bradley (2001). Reflections on the 25th anniversary of the individuals with disabilities education act. *Remedial and Special Education 22*(6), 324–334.

Kennedy-Lewis, B. L. and A. S. Murphy (2016). Listening to "frequent flyers": What persistently disciplined students have to say about being labeled as "bad". *Teachers College Record 118*(1), 1–40.

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

Lacoe, J. and M. P. Steinberg (2018). Rolling back zero tolerance: The effect of discipline policy reform on suspension usage and student outcomes. *Peabody Journal of Education 93*(2), 207–227.

Lazear, E. P. (2001). Educational production. *The Quarterly Journal of Economics 116*(3), 777–803.

Mittleman, J. (2018). A downward spiral? childhood suspension and the path to juvenile arrest. *Sociology of Education 91*(3), 183–204.

Morgan, P. L., G. Farkas, M. M. Hillemeier, R. Mattison, S. Maczuga, H. Li, and M. Cook (2015). Minorities are disproportionately underrepresented in special education: Longitudinal evidence across five disability conditions. *Educational Researcher 44*, 278–292.

Morgan, P. L., Y. Wang, A. D. Woods, Z. Mandel, G. Farkas, and M. M. Hillemeier (2019). Are u.s. schools discriminating when suspending students with disabilities? a best-evidence synthesis. *Exceptional Children 86*(1), 7–24.

Morgan, P. L., A. D. Woods, Y. Wang, G. Farkas, M. M. Hillemeier, and C. Mitchell (2023). Which students with disabilities are placed primarily outside of us elementary school general education classrooms? *Journal of Learning Disabilities 56*(3), 180–192.

Myderrizi, A. (2023). Teacher preparation, classroom structure, and learning outcomes for students with disabilities. EdWorkingPaper: 23-858. Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/yx96-2436.

Noltemeyer, A. L., R. M. Ward, and C. McLoughlin (2015). Relationship between school suspension and student outcomes: A meta-analysis. *School Psychology Review 44*(2), 224–240.

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

Pope, N. G. and G. W. Zuo (2021). Suspending suspensions: The education production consequences of school suspension policies.

Schwartz, A. E., B. G. Hopkins, and L. Stiefel (2021). The effects of special education on the academic performance of students with learning disabilities. *Journal of Policy Analysis and Management 40*(2), 480–520.

Skiba, R. J., M. I. Arredondo, and N. T. Williams (2014). More than a metaphor: The contribution of exclusionary discipline to a school-to-prison pipeline. *Equity and Excellence in Education 47*(4), 546–564.

Skiba, R. J., A. J. Artiles, E. B. Kozleski, D. J. Losen, and B. G. Harry (2016). Risks and consequences of oversimplifying educational inequities: A response to morgan et al. (2015). *Educational Researcher 45*(3), 221–225.

Sorensen, L. C., S. F. Bushway, and E. J. Gifford (2022). Getting tough? the effects of discretionary principal discipline on student outcomes. *Education Finance and Policy*.

Steinberg, M. P. and J. Lacoe (2016). What do we know about school discipline reform? *Education Next*. https://www.educationnext.org/what-do-we-know-about-school-discipline-reform-suspensions-expulsions/.

Steinberg, M. P. and J. Lacoe (2018a). The academic and behavioral consequences of discipline policy reform: Evidence from philadelphia. *35*.

Steinberg, M. P. and J. Lacoe (2018b). Reforming school discipline: School-level policy implementation and the consequences for suspended students and their peers. *American Journal of Education 125*(1), 29–77.

Stiefel, L., S. S. Fatima, J. R. Cimpian, and K. O'Hagan (2023). The role of school context in explaining racial disproportionality in special education. https://doi.org/10.26300/SGA6-A314.

Stiefel, L., M. Gottfried, M. Shiferaw, and A. Schwartz (2021). Is special education improving? case evidence from new york city. *Journal of Disability Policy Studies 32*(2), 95–107.

Theobald, R. J., D. D. Goldhaber, T. M. Gratz, and K. L. Holden (2019). Career and technical education, inclusion, and postsecondary outcomes for students with learning disabilities. *Journal of Learning Disabilities 52*(2), 109–119.

USDOE (2018). Individuals with disabilities education act; regulations part b, subpart a, sec 300.8 (c). https://sites.ed.gov/idea/regs/b/a/300.8/c.

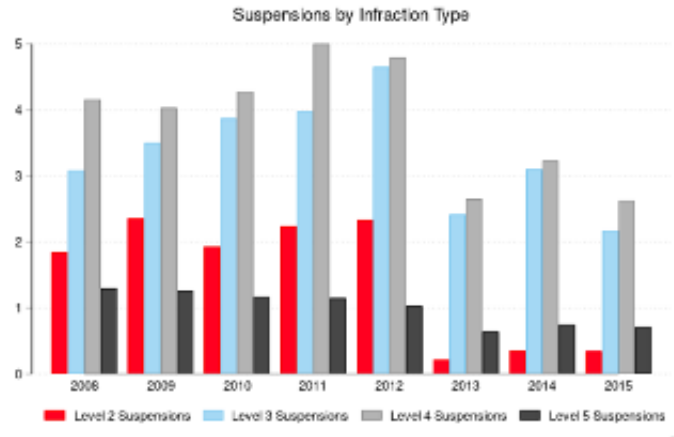
Wang, R. (2022). The impact of suspension reforms on discipline outcomes: Evidence from california high schools. *AERA Open 8*.

Welsh, R. O. and S. Little (2018a). Caste and control in schools: A systematic review of the pathways, rates and correlates of exclusion due to school discipline. *Children and Youth Services Review 94*, 315–339.

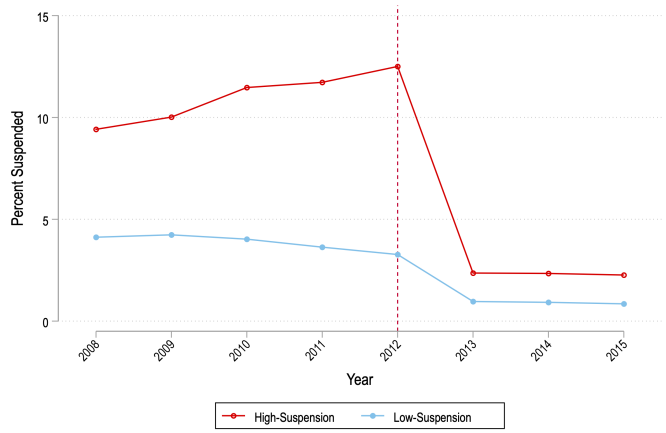
Welsh, R. O. and S. Little (2018b). The school discipline dilemma: A comprehensive review of disparities and alternative approaches. *American Educational Research Association 88*(5), 752–794.

Figure 1: Change in Infractions Associated with Suspensions

(a) Percent Suspended by Infraction Type



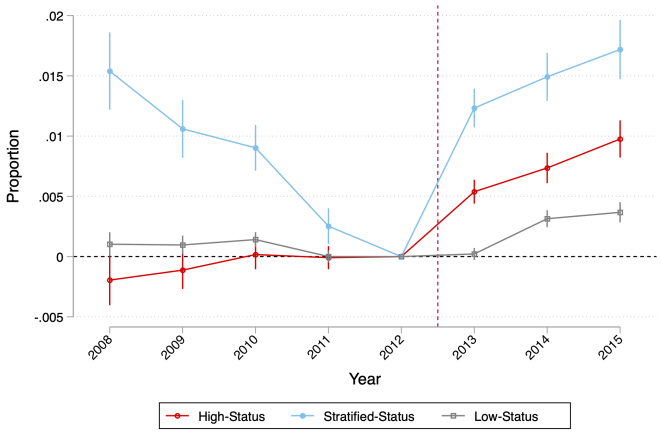
(b) Percent Suspended by Treatment Status



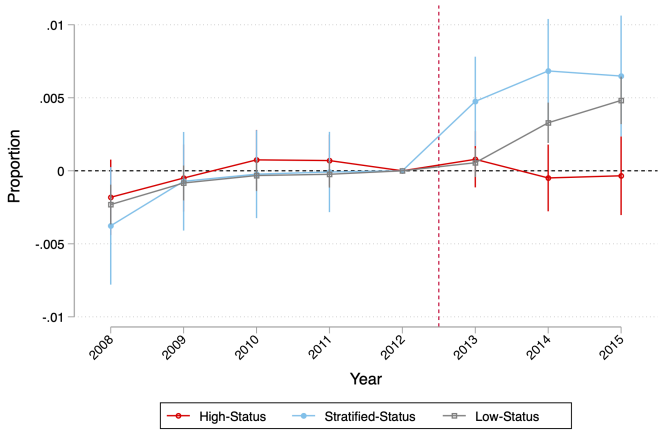
Notes. Panel A uses data from the Research Alliance for New York City Schools (RANYCS), which is hosted on a server that is separate from the data used for the rest of this paper. RANYCS data is not used for the remainder of the paper as their data on specific disability classification is not reliable pre-2012. As such, the sample for which the above estimates are derived may differ slightly from that used in the rest of this study. Estimates indicate the number of students per 100 that were suspended for a given infraction type for middle and high schools in NYCDOE. Panel B depicts the school-level suspension rates for above and below-median suspension schools pre- and post-ban using data hosted on the NYCDOE server.

Figure 2: Effects of the Suspension Ban on Classification Rate by Treatment Types

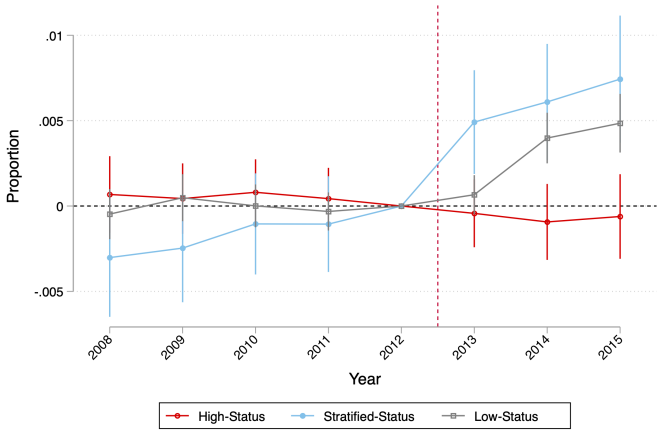
(a) Trends in Classification Pre- and Post-Ban



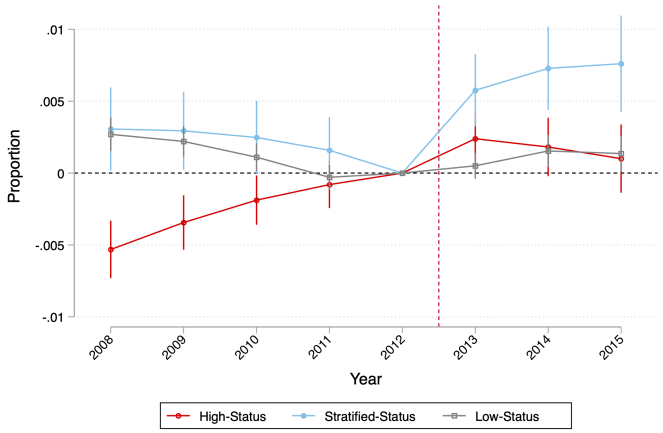
(b) Above−Below Median Suspension



(c) Black−Non-Black Students

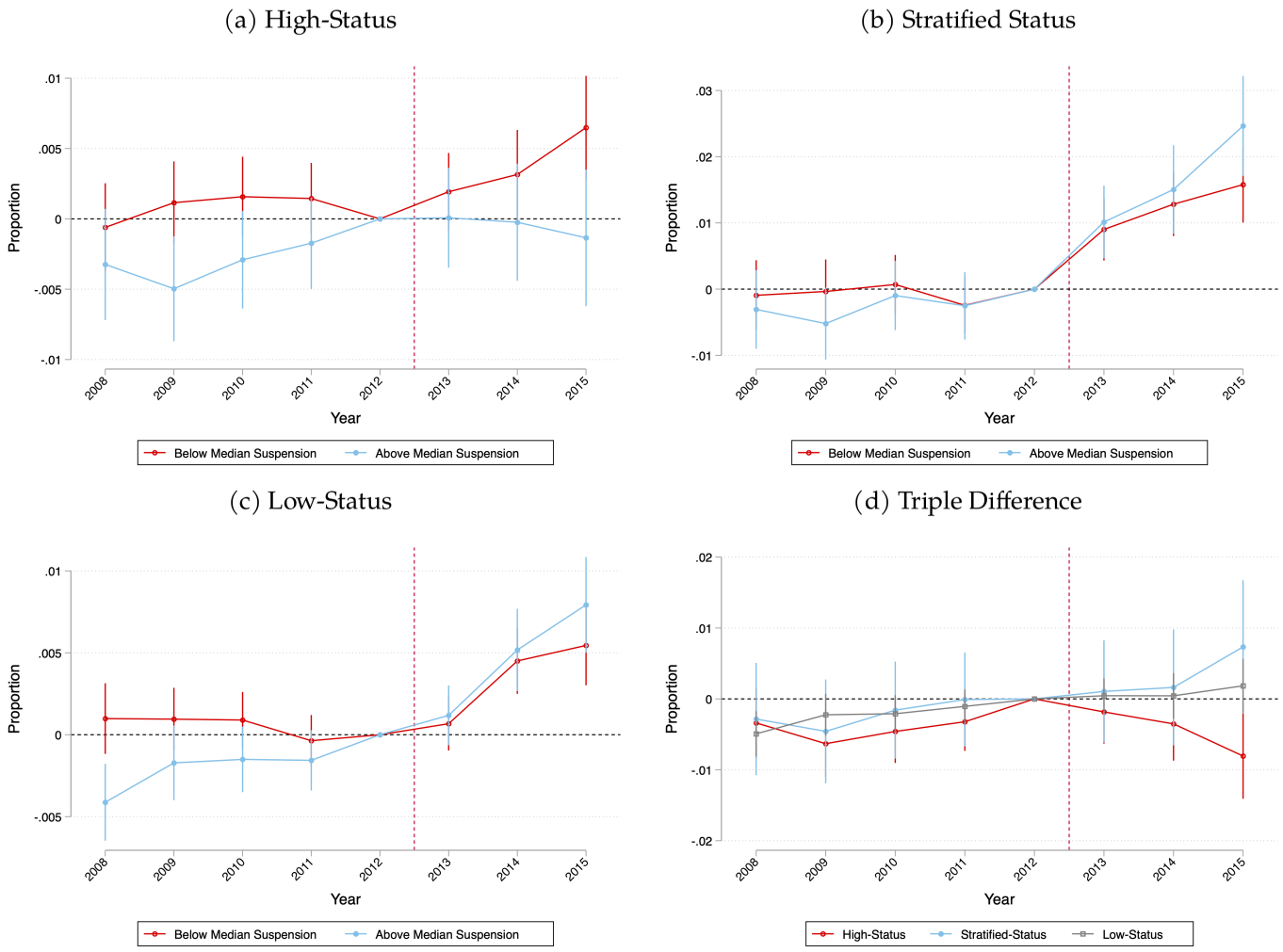


(d) Male−Female Students



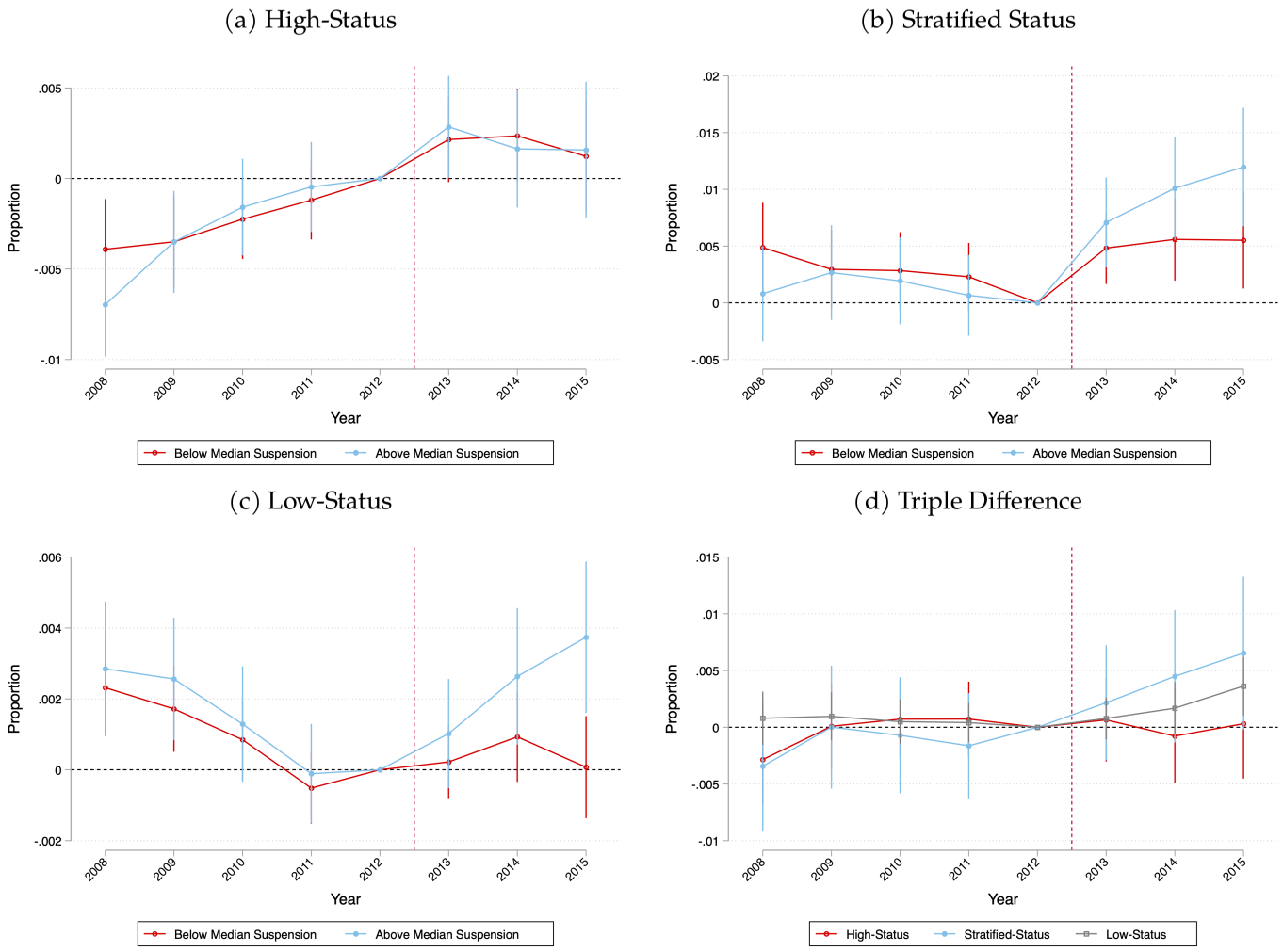
Notes.Estimates presented correspond to the trend over time disproportionality estimates in Tables 2 and 3, respectively. Estimates indicate the average change in classification rates relative to the baseline academic year (2011-12) for all middle and high school students. 95% confidence intervals are provided.

Figure 3: Black−White/Asian Students in High and Low Suspension Schools



Notes. Estimates in Panels A-C presented represent the Black-non-Black gap in classification within a high or low suspension school (above or below the 2011-12 district median suspension rate of 6 percent). Estimates in Panel D indicate the triple difference coefficient. 95% confidence intervals are displayed.

Figure 4: Male−Female Students in High and Low Suspension Schools



Notes. Estimates in Panels A-C presented represent the male-female gap in classification within a high or low suspension school (above or below the 2011-12 district median suspension rate of 6 percent). Estimates in Panel D indicate the triple difference coefficient. 95% confidence intervals are displayed.

Table 1: Descriptive Statistics

	Below	Above	Total
% of school SWD in pre-ban period	9.128	13.183	11.031
	(5.722)	(5.743)	(6.079)
% of school suspended in pre-ban period	3.913	10.731	7.112
	(2.674)	(4.908)	(5.165)
Average standardized math score	0.251	-0.238	0.021
	(0.949)	(0.842)	(0.933)
Average standardized reading score	0.180	-0.236	-0.015
	(0.977)	(0.851)	(0.944)
Years present	6.018	6.015	6.017
	(1.317)	(1.401)	(1.357)
FRPL Eligible (proportion)	0.585	0.670	0.625
English Language Learner (proportion)	0.073	0.100	0.086
Female (proportion)	0.503	0.484	0.494
Race (proportion)			
Black	0.234	0.377	0.301
Hispanic	0.362	0.443	0.400
White	0.182	0.079	0.134
Asian	0.217	0.097	0.161
Other	0.005	0.004	0.005
Grade			
6th grade	0.097	0.115	0.106
7th grade	0.123	0.150	0.136
8th grade	0.142	0.175	0.157
9th grade	0.205	0.201	0.203
10th grade	0.187	0.163	0.176
11th grade	0.133	0.107	0.121
12th grade	0.112	0.089	0.101
Disability Status (proportion)			
High-status classification	0.035	0.046	0.040
Stratified-status classification	0.082	0.115	0.098
Low-status classification	0.009	0.016	0.012
Hard classification	0.003	0.004	0.003
Observations	1,374,727	1,215,444	2,590,171
Schools	312	413	725

Note. Means and standard deviations (in parentheses) are displayed for continuous variables. Only means are displayed for binary variables. "Below" includes students in schools that had a below-median (6%) suspension rate in the year prior to the ban (2011-12). Unless otherwise indicated, values represent the means and standard deviations across the whole timeframe (2007-08 through 2014-15) for the complete sample, including District 75 schools.

Table 2: Trends in Classification Across Time

	Low-Status	**Stratified Status**	**High-Status**
2012 Baseline Proportion	*0.012*	*0.094*	*0.041*
2008	0.001*	0.015***	-0.002
	(0.001)	(0.002)	(0.001)
2009	0.001*	0.011***	-0.001
	(0.000)	(0.001)	(0.001)
2010	0.001***	0.009***	0.000
	(0.000)	(0.001)	(0.001)
2011	0.000	0.003***	-0.000
	(0.000)	(0.001)	(0.000)
2013	0.000	0.012***	0.005***
	(0.000)	(0.001)	(0.001)
2014	0.003***	0.015***	0.007***
	(0.000)	(0.001)	(0.001)
2015	0.004***	0.017***	0.010***
	(0.000)	(0.001)	(0.001)
Observations	2,590,171	2,590,171	2,590,171
Clusters	2,910	2,910	2,910

Notes. Standard errors in parentheses ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$) and are clustered at the school-grade level. All models use school-by-grade fixed effects and control for student gender, race, ELL status, FRPL eligibility, years present in NYCDOE, mean pre-ban ELA and math test scores, and a school's average pre-ban SWD classification rate interacted with a linear time trend.

Table 3: Difference-in-Difference Estimates

	Low-Status	Stratified Status	High-Status
Panel A. Above-Below Median Suspension			
2012 Baseline Difference	*0.007*	*0.031*	*0.012*
2013	0.000	0.011***	0.005***
	(0.000)	(0.001)	(0.001)
2014	0.002***	0.013***	0.008***
	(0.000)	(0.001)	(0.001)
2015	0.002***	0.015***	0.010***
	(0.000)	(0.001)	(0.001)
Above Median × 2013	0.001	0.005**	0.001
	(0.001)	(0.002)	(0.001)
Above Median × 2014	0.003***	0.007***	-0.000
	(0.001)	(0.002)	(0.001)
Above Median × 2015	0.005***	0.006**	-0.000
	(0.001)	(0.002)	(0.001)
Panel B. Black-Non-Black			
2012 Baseline Difference	*0.011*	*0.023*	*-0.003*
2013	-0.000	0.011***	0.005***
	(0.000)	(0.001)	(0.001)
2014	0.001***	0.013***	0.007***
	(0.000)	(0.001)	(0.001)
2015	0.001***	0.014***	0.010***
	(0.000)	(0.001)	(0.001)
Black × 2013	0.001	0.005**	-0.000
	(0.001)	(0.002)	(0.001)
Black × 2014	0.004***	0.006***	-0.001
	(0.001)	(0.002)	(0.001)
Black × 2015	0.005***	0.007***	-0.001
	(0.001)	(0.002)	(0.001)
Panel C. Male-Female			
2012 Baseline Difference	*0.012*	*0.036*	*0.032*
2013	-0.000	0.009***	0.004***
	(0.000)	(0.001)	(0.001)
2014	0.002***	0.010***	0.006***
	(0.000)	(0.001)	(0.001)
2015	0.002***	0.012***	0.009***
	(0.000)	(0.001)	(0.001)
Male × 2013	0.001	0.006***	0.002**
	(0.000)	(0.001)	(0.001)
Male × 2014	0.002**	0.007***	0.002
	(0.001)	(0.001)	(0.001)
Male × 2015	0.001*	0.008***	0.001
	(0.001)	(0.002)	(0.001)
Observations	2,590,171	2,590,171	2,590,171
Clusters	2,910	2,910	2,910

Notes. Standard errors in parentheses ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$) and are clustered at the school-grade level. All models use school-by-grade fixed effects. All models control for student gender, race, ELL status, FRPL eligibility, years present in NYCDOE, mean pre-ban ELA and math test scores, a school's average pre-ban suspension rate interacted with a linear time trend (except above/below), and a school's average pre-ban SWD classification rate interacted with a linear time trend. Above median indicates that a school's suspension rate in the 2011-12 AY was greater than the district median of 6%.

Table 4: Triple Difference Estimates: Male Students or Black Students in Above Median Suspension Schools

	Low Status	**Stratified Status**	**High Status**
Panel A. Male-Female			
2013 × Above Median × Male	0.001	0.002	0.001
	(0.001)	(0.003)	(0.002)
2014 × Above Median × Male	0.002	0.004	-0.001
	(0.001)	(0.003)	(0.002)
2015 × Above Median × Male	0.004**	0.007	0.000
	(0.001)	(0.003)	(0.002)
Observations	2,590,171	2,590,171	2,590,171
Clusters	2,910	2,910	2,910
Panel B. Black-White/Asian			
2013 × Above Median × Black	0.000	0.001	-0.002
	(0.001)	(0.004)	(0.002)
2014 × Above Median × Black	0.000	0.002	-0.004
	(0.002)	(0.004)	(0.003)
2015 × Above Median × Black	0.002	0.007	-0.008**
	(0.002)	(0.005)	(0.003)
Observations	1,542,579	1,542,579	1,542,579
Clusters	2,867	2,867	2,867

Notes. Standard errors in parentheses ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$) and are clustered at the school-grade level. All models use school-by-grade fixed effects and control for student gender, race, ELL status, FRPL eligibility, years present in NYCDOE, mean pre-ban ELA and math test scores, and a school's pre-ban SWD classification rate interacted with a linear time trend. Above median indicates that a school's suspension rate in the 2011-12 AY was greater than the district median of 6%. School by grade and year fixed effects are included in all models.

Table 5: Placebo Test

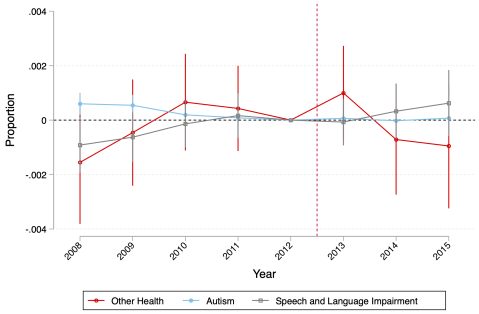
	Trends across Time	Above-Below	Black-Non-Black	Male-Female
2013	0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
2014	0.000	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
2015	0.000	0.001	0.001	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Group x 2013		-0.000	-0.000	0.000
		(0.000)	(0.000)	(0.000)
Group x 2014		-0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)
Group x 2015		-0.000	0.000	0.000
		(0.001)	(0.000)	(0.000)
Observations	2,590,171	2,590,171	2,590,171	2,590,171
Clusters	2,910	2,910	2,910	2,910

Notes. Standard errors in parentheses ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$) and are clustered at the school-grade level. All models use school-by-grade fixed effects and control for student gender, race, ELL status, FRPL eligibility, years present in NYCDOE, mean pre-ban ELA and math test scores, a school's pre-ban suspension rate interacted with a linear time trend (except above/below), and a school's pre-ban SWD classification rate interacted with a linear time trend. Above median indicates that a school's suspension rate in the 2011-12 AY was greater than the district median of 6%.

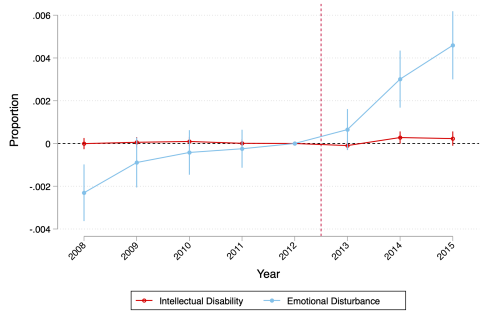
A Appendix

Figure A1: Estimates by Individual Disability Classification

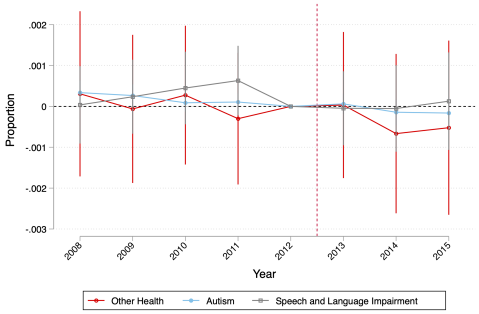
(a) Above−Below: High Status



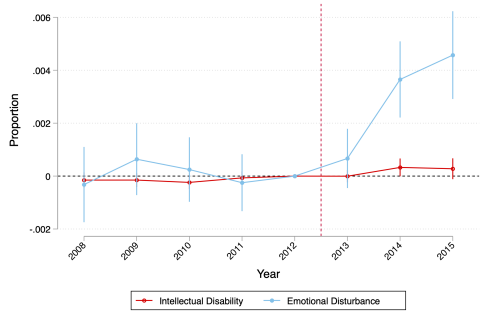
(b) Above−Below: Low Status



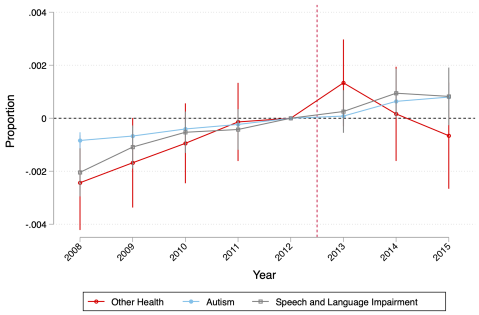
(c) Black−Non-Black: High Status



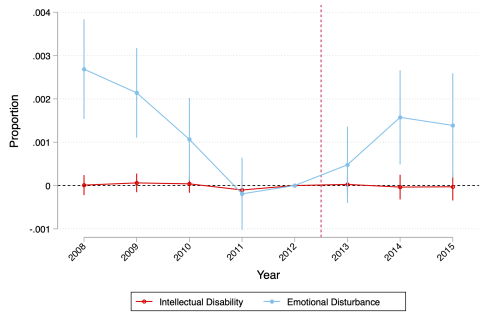
(d) Black−Non-Black: Low Status



(e) Male−Female: High Status

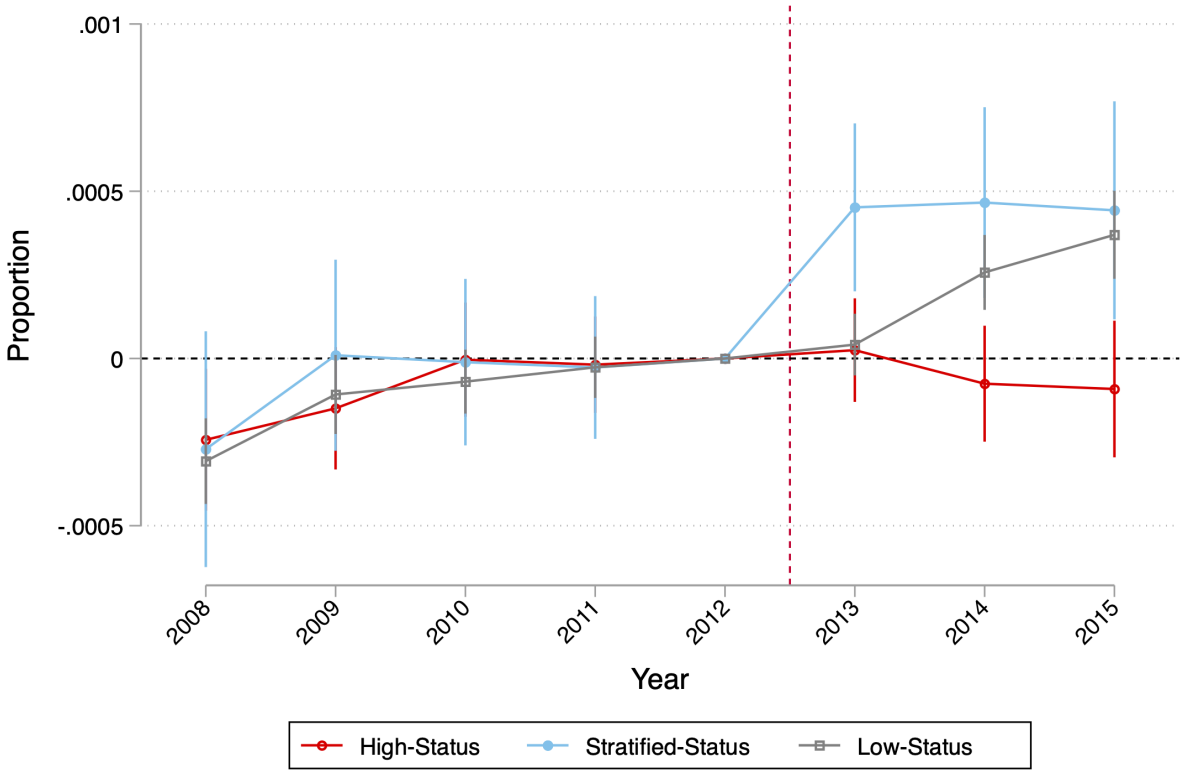


(f) Male−Female: Low Status



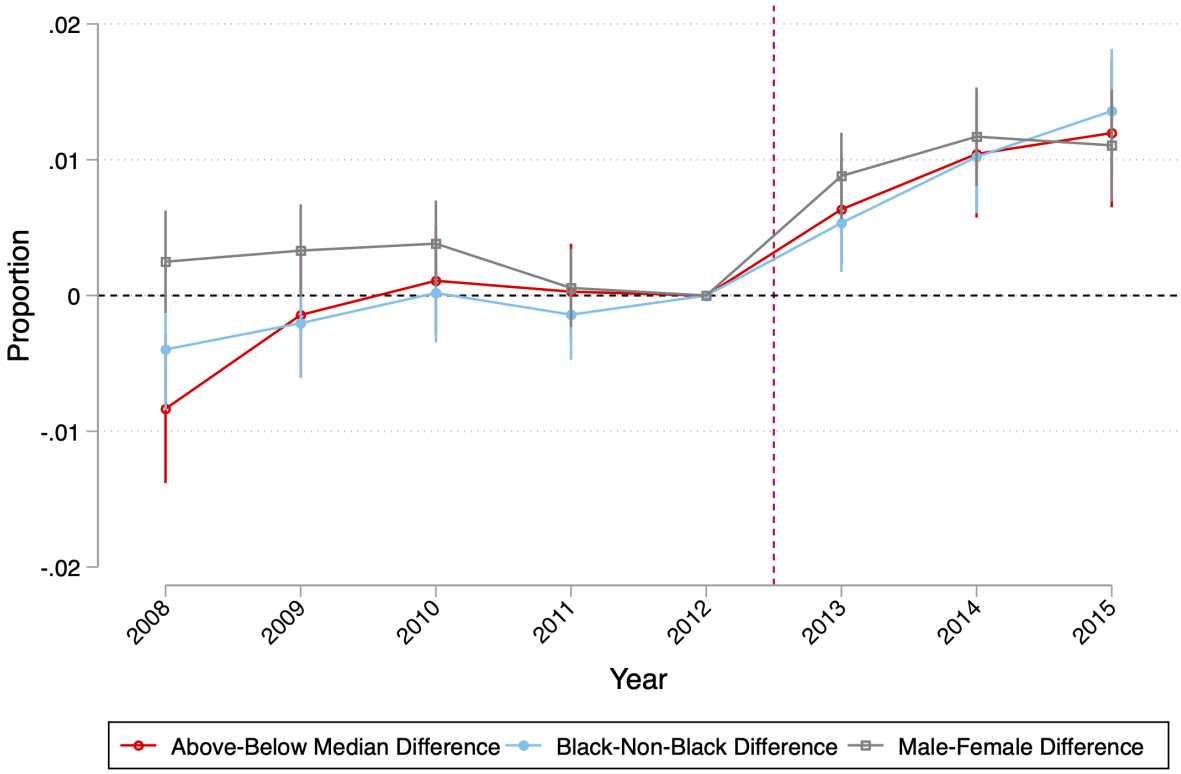
Notes. Estimates indicate the change over time relative to 2012 for trends over time, differences between Black and Non-Black students, and differences between male and female students for each individual classification type within low- and high-status groups. 95% confidence intervals are presented. All models use school-by-grade and year fixed effects and control for student gender, race, ELL status, FRPL eligibility, years present in NYCDOE, mean pre-ban ELA and math test scores, and a school's pre-ban SWD classification rate interacted with a linear time trend. Models estimating male-female or Black-Non-Black differences also control for a school's pre-ban suspension rate interacted with a linear time trend.

Figure A2: Continuous Treatment: Suspension Rate in 2011/12



Notes. Estimates indicate the change in disability classification relative to 2012 for a 1 standard deviation increase in a school's 2012 suspension rate. All models use school-by-grade and year fixed effects and control for student gender, race, ELL status, FRPL eligibility, years present in NYCDOE, mean pre-ban ELA and math test scores, and a school's pre-ban SWD classification rate interacted with a linear time trend. 95% confidence intervals are displayed.

Figure A3: Non-Imputed SWD Indicator for Any Classification



Notes. Estimates indicate the change in disability classification relative to 2012 for each treatment definition. All models use school-by-grade and year fixed effects and control for student gender, race, ELL status, FRPL eligibility, years present in NYCDOE, mean pre-ban ELA and math test scores, a school's pre-ban suspension rate (except Above-Below Median) interacted with a linear time trend, and a school's pre-ban SWD classification rate interacted with a linear time trend. 95% confidence intervals are displayed.

Table A1: 2012 Baseline Rates and Differences in Classification Rates

	Percent Classified in 2012		Ratio	Difference
	Below	Above		
High Status	3.57	4.73	1.33	1.17
Stratified Status	7.75	10.87	1.40	3.12
Low Status	0.76	1.43	1.89	0.68
Hard	0.28	0.32	1.15	0.04
	Non-Black	Black		
High Status	4.22	3.89	0.92	-0.32
Stratified Status	8.53	10.87	1.27	2.33
Low Status	0.75	1.85	2.46	1.10
Hard	0.31	0.28	0.92	-0.02
	Female	Male		
High Status	2.50	5.70	2.28	3.20
Stratified Status	7.38	11.02	1.49	3.64
Low Status	0.49	1.65	3.37	1.16
Hard	0.26	0.34	1.35	0.09

Notes. Estimates indicate the percent of students within each group that are classified with a specific disability. Ratio represents above/below; male/female; or Black/non-Black.

Table A2: Risk of Disability Classification and Self-Contained Services on Average and by Treatment Groups

	All Students		Disability Risk		Self-Contained Risk	
	% of Classifications	*% Self-Contained*				
Panel A. Above/Below Median						
			Below	*Above*	*Below*	*Above*
High-Status						
Autism	0.5	0.5	0.76	0.28	22.09	39.17
Other Health	19.87	21.99	19.80	19.93	29.73	30.76
Speech-Language Impairment	5.55	4.61	6.05	5.14	20.21	25.15
Stratified-Status						
Specific Learning Disability	62.9	56.58	63.07	62.76	24.37	24.84
Low-Status						
Emotional Disturbance	7.39	12.7	6.24	8.32	46.56	47.37
Intellectual Disability	0.38	1.09	0.35	0.40	84.65	74.52
Hard classification	2.15	1.71	2.40	1.94	21.23	22.27
Panel B. Black/Non-Black Students						
			Non-Black	*Black*	*Non-Black*	*Black*
High-Status						
Autism	0.5	0.5	0.62	0.27	26.20	32.98
Other Health	19.87	21.99	21.77	16.30	29.63	32.01
Speech-Language Impairment	5.55	4.61	5.92	4.84	20.75	27.35
Stratified-Status						
Specific Learning Disability	62.9	56.58	62.21	64.21	23.95	25.88
Low-Status						
Emotional Disturbance	7.39	12.7	5.53	10.90	46.60	47.51
Intellectual Disability	0.38	1.09	0.32	0.48	78.66	78.81
Hard classification	2.15	1.71	2.35	1.76	20.46	24.99
Panel C. Male/Female Students						
			Female	*Male*	*Female*	*Male*
High-Status						
Autism	0.5	0.5	0.15	0.69	35.55	26.51
Other Health	19.87	21.99	18.31	20.74	28.46	31.20
Speech-Language Impairment	5.55	4.61	3.46	6.70	20.58	23.36
Stratified-Status						
Specific Learning Disability	62.9	56.58	69.25	59.39	21.07	26.92
Low-Status						
Emotional Disturbance	7.39	12.7	4.56	8.95	42.16	48.44
Intellectual Disability	0.38	1.09	0.45	0.34	76.48	80.36
Hard classification	2.15	1.71	2.51	1.95	19.92	23.05

Notes. % of Classifications indicates the percent of all disability classifications that are associated with a given classification type. % Self-Contained indicates the percent of all students served in self-contained settings that have a given disability classification. Disability risk indicates the percent of SWDs within a group that are classified with a specific disability type. Self-Contained Risk indicates the percent of SWDs that have a specific disability type that are served in self-contained classrooms. Estimates use all available years of classification data.

Table A3: Robustness to a Shared Pathway to Success: Trends across Time and Difference-in-Difference Estimates

	Low-Status	Stratified Status	High-Status
Panel A. Trends across Time			
2013	0.000	0.012***	0.005***
	(0.000)	(0.001)	(0.001)
2014	0.002***	0.015***	0.007***
	(0.000)	(0.001)	(0.001)
2015	0.002***	0.017***	0.010***
	(0.000)	(0.001)	(0.001)
Panel B. Above-Below Median Suspension			
2013	0.000	0.010***	0.005***
	(0.000)	(0.001)	(0.001)
2014	0.001***	0.012***	0.007***
	(0.000)	(0.001)	(0.001)
2015	0.001**	0.015***	0.010***
	(0.000)	(0.002)	(0.001)
Above Median × 2013	-0.000	0.004**	0.001
	(0.000)	(0.002)	(0.001)
Above Median × 2014	0.001*	0.006***	-0.001
	(0.001)	(0.002)	(0.001)
Above Median × 2015	0.002**	0.006**	-0.000
	(0.001)	(0.002)	(0.001)
Panel C. Black-Non-Black			
2013	-0.000	0.010***	0.005***
	(0.000)	(0.001)	(0.001)
2014	0.001**	0.013***	0.007***
	(0.000)	(0.001)	(0.001)
2015	0.001**	0.014***	0.010***
	(0.000)	(0.001)	(0.001)
Black × 2013	0.000	0.005**	-0.001
	(0.001)	(0.002)	(0.001)
Black × 2014	0.003***	0.006***	-0.001
	(0.001)	(0.002)	(0.001)
Black × 2015	0.003***	0.008***	-0.001
	(0.001)	(0.002)	(0.001)
Panel D. Male-Female			
2013	-0.000	0.009***	0.004***
	(0.000)	(0.001)	(0.001)
2014	0.001***	0.010***	0.006***
	(0.000)	(0.001)	(0.001)
2015	0.002***	0.012***	0.009***
	(0.000)	(0.001)	(0.001)
Male × 2013	0.000	0.005***	0.002**
	(0.000)	(0.001)	(0.001)
Male × 2014	0.000	0.008***	0.002
	(0.000)	(0.001)	(0.001)
Male × 2015	-0.000	0.008***	0.001
	(0.001)	(0.002)	(0.001)
Observations	2,558,242	2,558,242	2,558,242
Clusters	2,910	2,910	2,910

Notes. Standard errors in parentheses ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$) and are clustered at the school-grade level. All models use school-by-grade fixed effects. All models control for student gender, race, ELL status, FRPL eligibility, years present in NYCDOE, mean pre-ban ELA and math test scores, a school's average pre-ban suspension rate interacted with a linear time trend (except above/below), and a school's average pre-ban SWD classification rate interacted with a linear time trend. Above/below controls for all but the average pre-ban suspension rate interacted with the time trend. Above median indicates that a school's suspension rate in the 2011-12 AY was greater than the district median of 6%. All models exclude students that were ever enrolled in a District 75 school.

Table A4: Triple Difference Estimates: Black-Non-Black in Above Median Suspension Schools

	Low Status	**Stratified Status**	**High Status**
2013 × Above Median × Black	0.000	0.000	-0.002
	(0.001)	(0.003)	(0.002)
2014 × Above Median × Black	-0.000	0.000	-0.004
	(0.001)	(0.003)	(0.002)
2015 × Above Median × Black	-0.000	0.000	-0.008**
	(0.002)	(0.004)	(0.003)
Observations	2,590,171	2,590,171	2,590,171
Clusters	2,910	2,910	2,910

Notes. Standard errors in parentheses ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$) and are clustered at the school-grade level. All models use school-by-grade fixed effects and control for student gender, race, ELL status, FRPL eligibility, years present in NYCDOE, mean pre-ban ELA and math test scores, and a school's pre-ban SWD classification rate interacted with a linear time trend. Above median indicates that a school's suspension rate in the 2011-12 AY was greater than the district median of 6%. School by grade and year fixed effects are included in all models.

Table A5: Correlation Coefficients between Measures of Pre-Policy Test Scores

	Mean Math	Last Available Math	Mean Reading
Last Available Math	0.94		
Mean Reading	0.74	0.68	
Last Available Reading	0.71	0.66	0.94

Notes. Estimates reflect the correlation coefficients between the primary test score measures, which are the mean of all available pre-policy middle school test scores, and a student's last available test score prior to the policy change.