ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 24-902

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

Jo Al Khafaji-King
New York University

Across the United States, suspension bans have become a popular policy response to address excessive and inequitable use of suspension in schools. However, there is little research that examines what strategies school staff employ when suspension is no longer permitted. I examine the effect of New York City's suspension ban on the use of a potential unintended substitute for suspension: special education classification. Using a dosage difference-in-differences strategy, I find that the ban induced an increase in disability classifications at high risk for classroom exclusion. I show that, on average, students with these classifications in schools with high pre-policy reliance on suspension experienced large declines in test scores, whereas general education students experienced slight test score improvements. Notably, I show that these declines are not due to new, ban-induced classifications actively harming student achievement. These results underscore the importance of considering unintended consequences and vulnerable groups when employing a seemingly "costless" and popular policy lever to reduce schools' reliance on suspension.

VERSION: September 2024

Suggested citation: Khafaji-King, Jo Al. (2024). Disability as Discipline? Effects of the New York City Suspension Ban on Identification of Students with Disabilities. (EdWorkingPaper: 24-902). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/yvwe-dq47

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

Jo Al Khafaji-King†

New York University

August 29, 2024

Abstract

Across the United States, suspension bans have become a popular policy response to address excessive and inequitable use of suspension in schools. However, there is little research that examines what strategies school staff employ when suspension is no longer permitted. I examine the effect of New York City's suspension ban on the use of a potential unintended substitute for suspension: special education classification. Using a dosage difference-in-differences strategy, I find that the ban induced an increase in disability classifications at high risk for classroom exclusion. I show that, on average, students with these classifications in schools with high pre-policy reliance on suspension experienced large declines in test scores, whereas general education students experienced slight test score improvements. Notably, I show that these declines are not due to new, ban-induced classifications actively harming student achievement. These results underscore the importance of considering unintended consequences and vulnerable groups when employing a seemingly "costless" and popular policy lever to reduce schools' reliance on suspension.

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

*This work benefited from funding from the Institute for Education Sciences sponsored Predoctoral Interdisciplinary Research Training fellowship (Grant R305B200010) and 2024 NAEd/Spencer Dissertation Fellowship. I am grateful to New York City Public Schools for providing access to the data used to support the findings in this paper. The opinions expressed are those of the myself and do not represent views of the IES, NAEd/Spencer, or NYCPS. I am indebted to Joseph Cimpian, Tatiana Homonoff, Leanna Stiefel, Luis Rodriguez, Riley Acton, Austin Smith, Lucy Sorensen, Marina Gorzig, Kaitlyn O'Hagan, Robin Caron, and Berta Bartoli for helpful comments and suggestions. I am grateful to participants 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 valuable feedback on this work. Any remaining errors are my own.

†**Contact:** jk7317@nyu.edu

For this reason he is sent to a special class... This is not because the backward class is the right place for him, but rather because it furnishes an easy means of disposing of a pupil who, through no fault of his own, is an unsatisfactory member of a regular grade.

Herbert A. Miller, 1916

1 Introduction

In recent years, education policymakers have restricted or banned the use of suspensions for minor misbehaviors (e.g., profane language or behavior, repeated tardiness, habitual disruption of the classroom, etc.). These policies were informed by extensive research that documents the detrimental impact of exclusionary discipline (e.g., suspension or expulsion) on student outcomes on average, and especially for marginalized groups (see Welsh and Little, 2018a,b). However, the nascent research base regarding the efficacy of suspension bans lacks consensus on how these policies impact student outcomes. Even more, no research has yet examined how student behavior is addressed in the wake of a suspension ban—what methods are school personnel using when suspension is no longer permitted? Understanding the potential alternative (and perhaps unintended) strategies that disciplinarians use subsequent to the implementation of these policies is important for determining the costs and benefits of such policies. For example, prior studies have referred to suspension bans as "costless" (Craig and Martin, 2023), but if these alternative strategies further stretch limited resources and result in worsened outcomes for specific subgroups, these policies are not necessarily equitable or efficient.

In this paper, I understand suspension bans as capacity constraints on school staff's permitted responses to student behavior, which induces a substitution to alternative responses to student behavior—some of which the policymaker may not intend. Guided by disability critical race (DisCrit) theory, I identify a plausible unintended substitute for suspension: special education classification. DisCrit theorists suggest that special education may be used as a mechanism of exclusion and behavior conformity (Kim et al., 2010; Annamma et al., 2013; Cruz et al., 2021). Specifically, these scholars have suggested that special education may function as a new and legalized form of segregation. I test this theory of disability classification as a mechanism of exclusion in the context of New York City's 2012 ban on suspension for minor offenses: I examine if the absence of one exclusionary strategy (suspension) may induce the use of another (special education). In addition, I estimate ban-induced changes to students' test scores and attendance and investigate potential mechanisms by which these changes occur.

To estimate these ban-induced changes, I leverage variation in a student's exposure to the suspension ban—some students attended schools that had a high pre-policy reliance on suspension while other students attended schools with low prior reliance on suspension. I employ a dosage difference-in-differences estimation strategy and define "more treated" students as those who enrolled in schools with above-median pre-ban suspension rates, which were forced to substitute to different strategies more than schools with low pre-ban suspension reliance. I find that the 2012 suspension ban induced an increase in disability classifications—by 2015, the proportion of students classified with disabilities increased by 1.1pp more in above-median suspension schools relative to below-median suspension schools, or 22% off the baseline.

I also explore heterogeneity within the "SWD" label. Guided by sociological frameworks that distinguish classifications based on their social status (Fish, 2019), I demonstrate that students with "low-status" disabilities (emotional disturbance or intellectual disabilities) are at high risk of exclusion. In contrast, students with "high-status" disabilities (autism, other health, and speech-language impairment) are not. Students with "stratified-status" disabilities (specific learning disabilities) vary in their risk of exclusion depending on student race. This suggests that low-status and, to an extent, stratified-status classifications may be close substitutes for suspension. Consistent with this theory, I find that low-status and stratified-status classifications drive the post-ban increases in disability classifications. The gap in low-status and stratified-status classification rates between students in above- and below-median suspension schools increased by 0.5pp (71% off the baseline) and 0.7pp (23% off the baseline) by 2015, respectively. I find no differential change in high-status classifications. Estimates using individual classifications instead of status groups produce similar patterns.

I also examine disproportionality in classification by student race and sex. Given that Black students and male students are also subjected to suspension at higher rates than non-Black or female students (i.e., are "more treated" by the ban), we should also expect that these students would experience increases in classifications. I find that both Black students and male students were more likely to receive a low- or stratified-status classification post-ban relative to non-Black and female students, respectively. Further, I examine patterns of classification for Black students in high-suspension schools and for male students in high-suspension schools using a triple differences strategy. I find that male students in high-suspension schools drive the increases in low-status and stratified-status classifications. There is minimal difference in classification rates between Black and non-Black students in high-suspension schools; however, this finding is likely due to extreme racial segregation in New York City Public Schools (NYCPS)—being a Black student in NYC is highly synonymous with being in a high-suspension school.

In addition to identifying plausible estimates of the effects of suspension bans on special education classification, I also estimate how SWDs on average were affected by the suspension ban relative to GENs. To do so, I compare outcomes for SWDs in high-suspension schools to outcomes for GENs in high-suspension schools using the same dosage difference-in-differences strategy. I find that SWDs classified with a low-status disability in high-suspension schools experienced declines in math test scores ($\sim$23% of a standard deviation) relative to their GEN peers, who experienced slight improvements ($\sim$4% of a standard deviation). For students with low-status classifications in high-suspension schools, their ELA test scores decreased by $\sim$25% of a standard deviation relative to their GEN peers. Estimates based on Jackson et al. (2021) imply that these reductions are equivalent to reducing per-student spending by $\sim$\$6000. I find negligible differential impacts of the suspension ban on attendance rates for SWDs. However, I do find that GENs in high-suspension schools experienced a 0.57pp reduction in attendance. While statistically significant, this estimate is practically quite small, translating to 1.03 fewer days of school. Notably, changes in test scores and attendance rates for students with stratified or high-status classifications—classifications that are perhaps less likely to be stigmatizing/exclusionary—are no different than changes for GENs.

Ban-induced changes in student outcomes may arise from different mechanisms. It may be the case that students who receive ban-induced classifications are harmed by these classifications. Simultaneously, it may also be the case that these test score declines are due to ban-induced changes in the classroom environment (e.g., increased disruption, reduced clarity regarding rules and norms, etc.). I test the role of newly identified, ban-induced classifications in contributing to these achievement declines, employing a student fixed effects intent-to-treat estimator, which avoids the issue of potentially endogenous student exit and re-entry to/from classification. I compare the effect of being classified in a high-suspension school to the effect of being classified in a low-suspension school and, further, compare the differences in effects relative to the last classification cohort before the ban (2011/12). I find that there is no difference in the effect of receiving a ban-induced (post-2012) low-status classification relative to non-ban-induced classification (pre-2012)—effect sizes remain similar across all years. Notably, the effect size of a stratified status classification on math scores post-2012 grew slightly more so for students classified in high-suspension schools than for those in low-suspension schools. If anything, students with ban-induced classifications may have benefited from services more so than students with non-ban-induced classifications.

Taken together, I suggest that the reduction in SWDs' test scores is likely *not* a byproduct of ban-induced classifications functioning differently than non-ban-induced classifications—ban-induced classifications functioned similarly or perhaps better than non-ban-induced classifications. Ruling out this mechanism suggests that perhaps suspension may have prohibited these students from receiving services and, even if classification was used as a mechanism of exclusion, it was preferable to suspension. As such, it may be the case that the reduction in test scores on average for SWDs with low-status disabilities is primarily attributable to ban-induced negative peer effects and/or rapid and unexpected changes to classroom norms, rules, and resources to which SWDs with low-status disabilities may be exceptionally sensitive. As such, future suspension policy must not only consider how disability classification may become more common (and, thus, require further resources needed to serve these students), but also how certain SWDs may not respond well to rapid changes in classroom norms and rules.

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 relatively nascent literature regarding the efficacy of suspension restrictions. I rely on a context (NYC) that had largely successful implementation without changes in other suspension types[1] and document the use of an understudied strategy of exclusion (special education classification). I also build on the strand of literature that addresses the effects of suspension reform on student's achievement outcomes. Prior studies have found mixed results—for example, Craig and Martin (2023), Karger and

[1] 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).

Komisarow (2024), and Cleveland (2023) find that suspension bans in different contexts improved student test scores in the aggregate, whereas Pope and Zuo (2023), Steinberg and Lacoe (2018b), and Anderson et al. (2019) find harmful effects on test scores or absenteeism on average, driven by negative peer effects or issues with implementation. I provide the first estimates of how suspension bans may differentially affect an exceptionally vulnerable group: students with disabilities.

Second, I contribute to the literature regarding theoretical and empirical links between suspension and disability classification. I draw from DisCrit to understand that these classifications may function 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 (suspension) may induce substitution to the other (special education). Further, I highlight the tension between the notion of disability classification as a mechanism of exclusion and the need to provide students with the services they need to succeed (Kim et al., 2010) and suggest that special education may simultaneously be used as a tool of segregation and support.

Lastly, my findings are relevant to the small, yet growing, quantitative literature on the effects of special education (O'Hagan and Stiefel, 2024). I provide the first estimates using a heterogeneous treatment robust estimator and show how suspension policy context slightly moderates the efficacy of these services in improving student outcomes. My findings generally align with the literature that suggests that special education, both broadly and for the marginal student, leads to increased educational attainment (Ballis and Heath, 2021b), improved test scores (Schwartz et al., 2021; Hwang, 2024), and reduced suspension likelihood (Hurwitz et al., 2021).[2] Notably, however, Black students may be adversely affected by misclassification (Ballis and Heath, 2021a). The generally positive effects of special education (O'Hagan and Stiefel, 2024) are in contrast to the well-documented phenomenon of suspension leading to worsened student outcomes (see, for example Noltemeyer et al., 2015; Welsh and Little, 2018b; Sorensen et al., 2022; Bacher-Hicks et al., 2019; Mittleman, 2018; Holt et al., 2022; Davison et al., 2021). As such, even if it is the case that

[2] Hwang (2024) uses a student fixed effects model and finds *increased* suspension likelihood for students with learning disabilities.

newly classified students are receiving these classifications with the intention to exclude, it seems that this strategy would be preferable to suspension.[3]

2 Existing Evidence on School Discipline Reform

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, 2018b). For example, suspending or expelling a student can lead to a lower likelihood of graduation (Sorensen et al., 2022); increased rates of incarceration (Bacher-Hicks 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). Further, 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; Elder et al., 2021). 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).

One strand of literature regarding suspension bans focuses on implementation fidelity. 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 suc-

[3] Of course, other alternative, recommended practices, such as restorative practices or progressive discipline are likely preferable to inducing inefficiencies in the provision of special education. I elaborate on other potential (and intended) strategies that may be used in the wake of the suspension ban in Section 3.

cessfully 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. Conversely, evidence from Lincove et al. (2024) suggests that this substitution to other suspension types did not occur in Maryland. Similarly, a concurrent evaluation by Craig and Martin (2023) shows that the New York City ban was met with fidelity from schools—the suspension rate for infractions targeted by the ban was met with near perfect compliance and the overall rate of suspensions for other infraction levels was also reduced, suggesting that the ban had spillover effects on reducing suspensions for higher level offenses.

A second strand of literature documents suspension bans' 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 improved on average for *all* students, which the authors attribute to improvements in school climate and, likely, the high compliance with the policy (Craig and Martin, 2023). Evidence from Massachusetts' efforts to reduce all suspensions presents similar results, with the reform improving ELA test scores, absenteeism, and dropout rates (Cleveland, 2023). Reforms affecting students in early grades in North Carolina (Karger and Komisarow, 2024) and Maryland (Lincove et al., 2024) successfully reduced suspension rates on average and for marginalized groups. Similarly, Rhode Island's reform also reduced suspension rates on average as well as suspension disparities (Craigie, 2022). 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 (2023) provide evidence from Los Angeles, where reductions in suspensions reduced test scores on average—concentrated improvements for high suspension risk students were offset by small, yet diffuse, impacts on low suspension risk peers.

3 Conceptual Framework

Under the Bloomberg administration, New York City Public Schools (NYCPS) 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.[4] The ban was announced in August of 2012, prior to the start of the 2012/13 academic year and well after the end of the 2011/12 academic year, leaving little room for teacher or administrator anticipation of the policy (Baker, 2012). The NYC suspension ban fundamentally restructured how school staff could address student behavior. Unless the ban itself drastically (and immediately) improved the behavior of students, teachers would still need a way to address 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; Welsh and Little, 2018a; Pope and Zuo, 2023). Put simply, a suspension ban functions as a capacity constraint, generating a necessary substitute for disciplinarians to turn to when suspension is prohibited.

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. Importantly, other alternative practices, such as restorative practices (RPs) and progressive discipline were encouraged to address this behavior; however, the training that teachers received likely varied dependent on the culture of the school or an individual teacher's discipline philosophy. Moreover, RPs, even with sufficient staff buy-in and successful implementation strategies, may take up to 3 years to shift attitudes toward punishment and upwards of 5 years to fundamentally change a school's culture (Fronius et al., 2016). As such, these *recommended* practices were not necessarily *implemented* practices.[5]

Other ways that teachers may respond to the ban-induced capacity constraint to address "non-compliant" and "disorderly" behavior also 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

[4] Charter schools were not required to comply with this policy.

[5] Craig and Martin (2023) also suggest that RPs were difficult to immediately adopt in the New York City context: "according to one teacher whose school used restorative interventions in the post-reform period, there were too many disciplinary issues to resolve all of them in such a resource-intensive way"(p. 17).

(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 (e.g., send a student to the principal's office). Each of these processes functions similarly to how a student may be classified with a disability. While teachers may not intend for their students to be classified, 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 in the number of students who 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 greater contact between students and behavioral teams.

In addition to the mechanical mechanism by which disability classification may become more common post-reform, the examination of this strategy as a potential unintended consequence of suspension reform is motivated by the theoretical and empirical parallels between disability and exclusionary discipline. Due to the relatively subjective nature of many classifications, their designation is based on academic 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 instead be perceived as a disability and met with services, or may be perceived as a student simply being "bad," resulting in suspension. Indeed, Barnard-Brak et al. (2023) estimate that 1 in 14 students who are expelled have an unserved disability. Further, 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).

Other empirical work also draws parallels between disability and suspension showing that both suspension and subjective disability classifications are influenced by similar student misbehavior (Bal et al., 2019; Huang, 2020; Cruz et al., 2021; Hurwitz et al., 2021). This is especially the case with lower-level behaviors (e.g., disorderly or non-compliant) that have more malleable definitions. Further, "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 from the GEN classroom (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 one method of exclusion (i.e., suspension), we should also expect the relationship between classification and exclusion to change in tandem, especially for the marginal student whose behavior is exceptionally subjective. 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).

Theoretically, DisCrit scholars have also argued that "dis/ability status justifies segregation and unequal treatment for students of color compared to their white counterparts" (Annamma et al., 2013). Ableism and racism reinforce each other to justify exclusion and labelling of minority students and further codify the need to "achieve normality" (i.e., to exclude racial minorities and dis/abled students) into societal and educational goals. At its worst, race becomes synonymous with ability such that an intellectual hierarchy becomes a racial hierarchy and vice versa. This reifies the use of exclusion through at least two channels: exclusionary discipline and exclusionary special education services. 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. In this vein, we can conceptualize special education classifications as potential mechanisms of exclusion, functioning to remove specific groups from the general education classroom.

Simultaneously, designating a student with a disability allows for procedural protections and services that are needed. Among other protections from exclusionary discipline, if a student is classified with a disability, they cannot be suspended for more than 10 days without a manifestation determination, which is an evaluation to determine if the behavior in question was a result of a student's disability. If the behavior was indeed a product of a disability, the student cannot be removed for more than 10 days. For students *without* disabilities, these protections are not afforded (Kim et al., 2010). Further, there is a strand of literature that documents the benefits to other student outcomes that disability classification provides (O'Hagan and Stiefel, 2024). While this evidence base is relatively small, existing evidence suggests that special education services improve math and ELA test scores (Schwartz et al., 2021), disciplinary outcomes (Hurwitz et al., 2021), and educational attainment (Ballis and Heath, 2021a).[6]

Importantly, the fact that special education classification provides procedural protections and tends to improve student outcomes does not necessarily preclude the fact that it may also be used as a mechanism of exclusion. The latter can still be true, especially when considering the counterfactual: if a student were not excluded via special education, what would be the alternative mechanism? Taken together, there arises a paradox: a student may be classified with a disability as a mechanism of exclusion but simultaneously and be afforded protections from exclusion and needed services upon classification. Regardless of whether students are receiving classifications as a mechanism of exclusion or as needed services, we would expect that classification should increase post-reform.

There is also substantial heterogeneity within the "SWD" label. Some classifications carry higher stigma and greater exclusion from the general education classroom, whereas others may carry a positive connotation and more inclusive services. Fish (2019) proposes a framework for understanding differences across classification types, grouping classifications into "low-status", "stratified status", and "high-status" based on their degree of exclusion and stigma associated with each label.[7] Low-status classifications are stigmatized as being 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; Anderson, 2021; Theobald et al., 2019). Moreover, Black students are persistently over-represented in low-status classifications, whereas non-Black students are over-represented in high-status classifications (Blanchett, 2010).[8,9] Stratified-status

[6]As noted in O'Hagan and Stiefel (2024), these studies do not necessarily imply causality.

[7]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.

[8]This is also observed in my data. See Tables A1 and A2 for racial differences in disability classification.

[9]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). However, there are unresolved debates regarding disproportionality estimates that control for these characteristics (Collins et al., 2016; Skiba et al., 2016): 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.

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)).

Given the heterogeneity within the SWD label at baseline, there is likely heterogeneity in the types of classifications that are expected to increase post-ban. 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, classify a student with a disability. If this hypothesis holds true, we should expect that classifications, especially those with greater inclusion and less stigma (i.e., high-status and, to an extent, stratified status classifications), 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 more likely to be included in the classroom. Conversely, if increases in classifications are concentrated primarily among low-status classifications, this would suggest that these classifications may have been used as a mechanism of exclusion.

Amidst increases in disability classifications, it is unclear how these students would be affected. If these classifications are functioning purely as a mechanism of exclusion, we may expect a student's outcomes may be harmed if they do not benefit from certain services and are simply labeled with a disability classification. This may be especially salient for low-status classifications. It may also be the case that students are now receiving needed services instead of being met immediately with suspension. In this case, these students on the margin of service reception likely benefit from being classified (and also being less likely to be suspended).

More generally, we may expect that SWDs on average (both those newly and previously classified) experience a differential impact of the suspension ban relative to their GEN peers. Given that SWDs at baseline are more likely to be suspended relative to GENs, a ban on suspensions may benefit SWDs' achievement insofar as they benefit from increased class time and a greater connection to their school, peers, and teachers. However, SWDs may also be exceptionally sensitive to changes in the schooling environment (O'Hagan et al., 2023). Further, a systematic review from Maciver et al. (2019) suggests that consistent and clear routines, rules, and norms are essential for

SWDs' participation in the classroom, broadly. These rules are especially salient for students with emotional behavioral disorders (EBD), which is the most common low-status classification. Simpson et al. (2011) highlights nine essential environmental supports for EBD students, which include not only clear and consistent rules, but also uniform feedback when rules/norms are violated as well as reducing student exposure to disruptive environments. In the context of a suspension ban, it is likely the case that changes to the rules and norms (even if they are presumably more inclusive) may be a difficult adjustment for students with an EBD classification. Further, these students may be especially sensitive to potential increases in disruption induced by the suspension ban. These potential outcomes may apply to both students already classified with a disability prior to the suspension ban, as well as those students who are newly classified.

4 Data

4.1 Administrative Data

I use student-level, administrative data provided by NYCPS 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,[10] demographic characteristics (race/ethnicity, sex, FRPL status, ELL status), whether a student was classified as a SWD, and their specific classification, if applicable. I include all students in middle and high schools in New York City, but exclude charter schools from the sample.[11] Ultimately, the sample covers over 500,000 unique students in over 700 middle and high schools throughout the city.

NYCPS 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 classified in 2012

[10] 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.

[11] 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 5 for further discussion of this limitation.

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.[12]

4.2 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 NYCPS 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 (deafness, blindness, deaf-blindness, orthopedic impairment, traumatic brain injury, and hearing impairment). 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-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.

[12]These imputations do not drive my results. I conduct a robustness check in FigureA4 that uses the non-imputed SWD indicator as my outcome, finding that my results for all specifications align.

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 primary results and conduct robustness checks using these individual classification types (Figure A5).

4.3 Examining Trends in Treatment and Outcomes across Time

At a baseline, it must be the case that the ban was implemented with fidelity and substantially reduced the use of suspension. Figure 1 presents raw suspension rates for schools with above-median and below-median suspension rates. For schools with above-median 2012 suspension rates (above 6 percent), 12.5 percent of students were suspended in 2012, whereas in below-median suspension schools, only 3.3 percent of students were suspended. In the year immediately after the ban, suspension rates in above-median suspension schools dropped by 10.1 percentage points (pp) to an average of 2.4 percent of students being suspended on average. Below-median suspension schools only experienced a 2.3pp drop in suspensions, with an average of 1 percent of students being suspended in these schools in 2013. This sharp reduction in suspension rates—which is significantly larger for above-median suspension schools—is the key source of identifying variation in my empirical strategy.[13]

Table 2 displays regression adjusted changes in classification rates for all students relative to 2011/12.[14] On average, the proportion of students classified with subjective disabilities increased post-ban. Stratified-status and high-status classifications drive these increases, with stratified-status classifications increasing by 1.7pp 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

[13] This drop in suspension rates largely aligns with that documented in Craig and Martin (2023).

[14] This is estimated using an interrupted time series model with disability classification as the outcome variable using school-grade fixed effects and year indicators with 2012 being the omitted year. As such, these coefficients indicate the change in classification rates on average relative to 2012. 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, as well as a linear time trend interacted with school-level average pre-ban share of students with disabilities. I do not interpret these estimates as causal.

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.

5 Empirical Strategy: Ban-Induced Changes in Classification Behavior

5.1 Event Study Difference-in-Differences

To identify the causal effect of the ban, I rely on 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). As shown in Figure 1, there is substantial variation in the dosage of the suspension ban "treatment." As such, I estimate dosage difference-in-difference models where treatment status 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.

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. As such, I rely on suspension data aggregated to the school-level to determine treatment status. This means that there is likely noise in treatment assignment. For example, a school could be identified as treated, but its suspensions consist of only Level 3 infractions. This school would be relatively unaffected by the ban, but would be identified as treated due to its high reliance on suspension on average. Therefore, the estimates presented are attenuated and are likely a lower bound for effect sizes.

I estimate the following equation:

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

where α_{sg} is a school-grade fixed effect; X_{isgt} are time-varying student characteristics; $\text{SWD}_{isgt,j}$ is an indicator for being a student i in school-grade sg in academic year t with a disability classification of type j where $j = \{\text{low, stratified, high}\}$; $\rho_{k,j}$ are the coefficients of interest, where "High Risk" indicates attending a school with above-median 2012 suspension rates[15]; ϵ_{isgt} is an idiosyncratic error term with the usual properties; X_{st} are time-varying school characteristics and include a linear time trend interacted with a school's pre-ban SWD classification rate. The coefficients $\rho_{k,j}$ can be interpreted as the difference-in-difference estimate—that is, the difference in the classification rates between high- and low-suspension schools, relative to their 2012 difference.[16]

In addition to some schools being more subject to treatment than others, certain demographic groups are also more treated. As such, I also provide estimates of changes in racial and gender disproportionality in disability classification, modifying equation 5.1 to one in which "High Risk" is either an indicator for a student being Black or an indicator for a student being male. To an extent, these estimates can also be thought of as dosage treatments, insofar as Black and male students are over-represented in suspension and, thus, more affected by the ban regardless of the school that they attend. As such, I compare changes in classification rates for Black students relative to their non-Black peers and effects for male students relative to their female peers.[17]

5.2 Event Study Difference-in-Differences: Moderation by Race and Sex

Given the coarseness of using school-level suspension rates to define my primary treatment, I aim to further pinpoint the impact of the suspension ban by relying on the fact that *within* schools certain demographics (Black students and male students) are over-represented in suspension rates. This allows me to examine within-school variation in treatment intensity, again, so long as being Black or being male increased the likelihood of suspension in the pre-ban period. I estimate a fully interacted model of the following form:

$$\text{SWD}_{isgt,j} = \beta + \alpha_{\text{sg}} + \sum_{k \neq 2012}^{2015} \rho_{k,j}[\mathbb{1}(t=k) \times \mathbb{1}(\text{High Suspension}_s) \times \mathbb{1}(\text{Demographic}_{isg,c})] + X_{\text{isgt}}\phi + X_{\text{st}}\eta + \epsilon_{\text{isgt}}$$

where $\text{Demographic}_{i,c}$ is a time-invariant indicator variable for characteristic c of student i in

[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). I also define treatment as being in the second or third tercile of 2011/12 suspension rates.

[16] 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).

[17] 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 students' disproportionality in discipline.

school-grade sg where $c = \{\text{Black, male}\}$. The coefficients of interest are $\rho_{k,j}$ and indicate the differential change in classification rates for being a high suspension risk demographic within a high-suspension school relative to a low suspension risk demographic in a high-suspension school. All other parameters retain the same meaning as in equation 5.1.

6 Results: Ban-Induced Changes in Classification Behavior

6.1 Effect on SWD Identification

Table 3 and Figure 2 present estimates for the differential effect of the suspension ban between schools with high and low suspension reliance. Relative to low-suspension schools in 2012, students in high-suspension schools in 2015 were 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 increased by 0.7pp ($p < 0.01$) for stratified-status classifications by 2015 (∼23 percent from the baseline difference of 3.1 percent) and 0.5pp ($p < 0.001$) 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 were 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 increased 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, induced to classifying students with low- and stratified-status disabilities.

6.2 Treatment and Disproportionality by Race and Sex

Next, I estimate specifications in which "High Risk" is either an indicator for a student being Black or an indicator for a student being male. These estimates can be thought of as a different definition of treatment given that Black and male students are more likely to be suspended.[18] In ad-

[18]This is true throughout the literature, as well as in the NYC context as demonstrated by Craig and Martin (2023) and (Rodriguez and Welsh, 2022).

dition 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 4, Panel A and Figure 3, Panel A 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 ($p < 0.001$) for stratified-status classifications and by 0.5pp ($p < 0.001$) for low-status classifications, approximately 30 and 45 percent from their respective baseline differences.[19] 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 comparing male students to female students in Table 4, Panel B and Figure 3, Panel B. 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 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.

6.3 Moderation by Race and Sex

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. While there is greater gender balance across above- and below-median suspension schools, there is a concentration of Black students in above-median schools.[20]

[19] Figure A2 presents estimates comparing Black students to White/Asian students.

[20] 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—essentially making being in a high-suspension school synonymous with "Black"—as well as the variability of the comparison group

Estimates for the triple difference by school type and race are presented in Table 5, Panel A and Figure 5. These estimates suggest that Black students in low-suspension schools are increasingly more likely than non-Black students to be classified with a high-status disability, whereas Black students in a high-suspension schools experience a decrease in high-status classifications relative to their non-Black peers. For schools with less punitive environments, it appears as though students who are over-represented in suspension begin to receive services. However, over-represented students, on average, are simultaneously more likely to receive low-status classifications.

To further investigate whether the effects observed are due to special education being a substitute for suspension, I estimate the ban's effect on males in above-median suspension schools. Figure 4 and Table 5, Panel B present these results. The difference in classification rates between males and females grows more in high-suspension schools than in 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 who are more affected by the suspension ban.

6.4 Robustness to Identifying Assumptions

6.4.1 Parallel Trends

For the event study difference-in-differences model to uncover causal effects of the ban, the primary assumption is that prior to the policy change, the treatment and comparison groups were trending similarly in their classification behavior, that is, these groups are plausible counterfactuals for each other. This implicitly assumes that, if but for the ban, these two groups would have continued on this same, parallel trend. Importantly, this does not necessitate that these groups have similar baseline *levels* of SWD classification, instead, that their trends prior to the ban were the

composition makes these estimates difficult to interpret, even with school-grade fixed effects. As such, I also 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). Appendix Figure A3 and Table A4 present these results between Black and White/Asian students. These estimates suggest that Black students in above-median suspension schools are more likely to receive a low- or stratified-status classification, but these results are not statistically significant at conventional levels.

same. While I cannot directly observe the counterfactual world in which high-suspension schools were not affected by the suspension ban, I can test for parallel pre-trends, which can be observed in each figure with point estimates for the 3 years leading up to the ban being no different from zero.

6.4.2 Falsification Test

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 falsification 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 the basic trends in classification and all primary models: high- and low-suspension treatment, Black vs. non-Black students, and male vs. female students. Table 6 presents these estimates. For all models, there is no change in hard classifications post-ban. 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 addressed in a school. Additionally, when estimating the impact of the ban by specific disability type (Figure A5), I find that emotional disturbance drives the estimates for low-status classifications, with little change 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 A5, Panels B, D, and F.

6.4.3 Alternative Treatment Specifications

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 A1, Panel A). I also provide an alternative measure of treatment, using terciles of pre-ban suspension rates to define a non-parametric gradient of treatment as in Sun et al. (2024). I find that these results are as expected—estimates from comparing schools in the second tercile of suspension to schools in the first tercile are slightly smaller in magnitude than those from my primary specification. Estimates from comparing third tercile suspension schools to first tercile suspension schools are slightly larger in magnitude. Results using these measures of treatment can be found in Figure A1 Panels B and C. 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 NYCPS. I find the same patterns, with classifications increasing more so for more-treated students than their respective counterfactuals (Figure A4).

6.4.4 Addressing Contemporaneous Policy Changes

A potential threat to the validity of my estimates 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 (Stiefel et al., 2021). Generally, this reform did not change *who* was classified but, instead *how* students received services. To ensure that the results presented are not due to aSPtS, I conduct a variety of robustness checks. I first exclude students who ever attended a District 75 school (those most affected by the inclusion initiative) prior to the ban.[21] Estimates using this restricted sample produce nearly identical estimates as the main results, albeit slightly attenuated, as expected.[22] All estimates are presented in Table A3.

Nonetheless, one may suggest that the introduction of aSPtS, while only intended to affect those students already classified, may affect classification behavior. For example, it may be the case that in special education programs where inclusion becomes more salient that parents and teachers are more aware of the resources that may be provided to their students. As such, they may be

[21] This excludes approximately 32,000 observations.

[22] 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.

more likely to advocate for services, resulting in an increase in service provision. Importantly, this phenomenon would only be a threat to validity if the inclusion initiative affected students in a way that also co-varied with whether a student attended a high- or low-suspension school. I test this hypothesis in two ways: first, I estimate whether there were differential decreases in the percent of SWDs served in self-contained classrooms between more affected and less affected groups (Tables 7); second, I define a new measure of "treatment" as being in a school with above-median baseline shares of SWDs served in self-contained settings (i.e., having high saliency in service provision) and re-estimate my primary equation using this treatment definition (Figure A6). Reassuringly, I find that there are no differences in movements out of self-contained settings between above- and below-median suspension schools or between male and female students.[23] Additionally, there is no differential change in classification behavior for those schools more affected by aSPtS (Figure A6).[24] Taken together, this evidence suggests the previously documented effects on classification behavior are largely attributable to the suspension ban, not aSPtS.[25]

6.4.5 Mobility and Attrition Analyses

Lastly, for my estimates to be treated as causal, it must be the case that the suspension ban did not influence other outcomes that would be correlated with classification rates. For example, one may be concerned that the result of increased disability classification is not due to the suspension ban, but instead due to differential attrition and mobility among students in above- and below-median suspension schools. I test the validity of this assumption by conducting balance tests, using racial composition, FRPL composition, sex composition, and non-structural movers, as my outcome variables. I provide the relevant estimates for these results in Table A6, which demonstrate that my primary results are due to the suspension ban, not due to ban-induced sorting across schools nor due to sample attrition. The one exception is a slight increase (0.4pp) in

[23] The exception is that Black students were more likely to move out of self-contained settings relative to non-Black students; however, as shown in Table 8, Black students in high-suspension schools were not more likely to move out of self-contained settings.

[24] There is evidence of pre-trends in these estimates, so these estimates should not be interpreted as completely causal, but instead as suggestive evidence that aSPtS had limited influence on classification behavior.

[25] To further show the lack of correlation between these two policy changes, I also provide a two-way contingency table that shows the overlap between students in schools with above-median movement out of self-contained settings and those schools with above-median suspension rates—the percent of students in schools with high levels of exclusion do not overlap tremendously with schools that have high levels of suspension rates(Table A5, Panel A). Similarly, there is little difference in baseline suspension rates between schools that have high or low baseline rates of SWDs served in self-contained settings, that is, minimal co-variance in treatment status (Table A5, Panel B).

non-structural movers—that is, students who switch/leave a school in a non-graduating grade (8 or 12)—in above-median suspension schools. I re-estimate my primary specification excluding all non-structural movers and find my results are unchanged (Table A7). Further, I examine patterns in sample attrition and show that while there is increased attrition in above-median suspension schools, the characteristics of these attritors are no different than those students who stay and are no more or less likely to be SWDs (Table A8).

7 Empirical Strategy: Effects of the Ban on Students with Disabilities

Given the documented increase in ban-induced disability classifications, specifically low- and stratified-status classifications, it is of interest to examine how SWDs may have been differentially affected by the suspension. It may be the case that the suspension ban reduced the likelihood that SWDs would be suspended, leading to increased class time and improvements in test scores and attendance. Conversely, it may be the case that SWDs respond poorly to rapid changes in classroom norms, rules, and, to an extent, potential increased disruption in the classroom. In this way, the impact of the ban on SWDs on average is ambiguous.

Further, it is also of interest to determine if ban-induced classifications are functioning more so as a mechanism of exclusion or as provision of needed services. Indeed, as noted in Kim et al. (2010), there is a tension between ensuring that special education more broadly is not being used as a means of exclusion while simultaneously ensuring that under-resourced students receive the services they need. If these classifications are being used as mechanisms of exclusion, we may expect that SWDs are negatively impacted by receiving these classifications. However, if these are students on the margin who simply needed services, we may expect that these students' outcomes would improve at best and, at worst, not be impacted.

7.1 Examining Differential Impacts of the Ban for SWDs on Average

To estimate the differential impacts of the ban for SWDs relative to GENs, I modify equation 5.1, estimating a specification of the following form:

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

where $Y_{isgt,j}$ is the outcome of interest (i.e., math and ELA test scores, attendance rates), α_{sg} is a school-grade fixed effect, High Risk_s indicates whether a school has an above-median pre-ban suspension rate[26], $\text{SWD}_{isgt,j}$ is an indicator for being classified with a disability of type j, where $j = \{\text{low, stratified, high}\}$, and $\rho_{k,j}$ are the coefficients of interest, which indicate the difference in outcomes for SWDs relative to GENs in high-suspension schools relative to the difference in outcomes between those groups in 2012.[27] I include student-level controls for race, sex, ELL and FRPL status, and years in the analytic sample ($\text{SWD}_{isgt,j}$). I also include an interaction term between a school's pre-ban proportion of SWDs interacted with a linear time trend (X_{st}).[28,29]

The estimates of the parameters $\rho_{k,j}$ include effects from multiple potential channels. SWD outcomes may differ due to the changes in the type of student that is being classified (i.e., ban-induced classifications) and they may change due to SWDs responding to ban-induced changes in norms, rules, and classroom practices differently than their GEN peers. As such, $\rho_{k,j}$ comprise both this response to environmental changes as well as the change in classification behavior. Further, it may be the case that above-median and below-median suspension schools have differences in their ability to accommodate the influx of students from self-contained classrooms due to aSPtS, even if that influx was of the same magnitude (see Table 7). To mitigate the direct contaminating influence of aSPtS on my estimates, I restrict my sample to SWDs who are only ever served in inclusive settings such that changes in a SWD's post-2012 outcomes are not attributable to their personal movement from a self-contained setting to an inclusive setting. This does not, however, rule out the potential peer effects of increased inclusion, especially if above- and below-median suspension schools have different capacities to handle otherwise uniform increases in inclusion.[30]

[26]Because test scores and attendance rates outcomes apply to different analytic samples (i.e., test scores are only available for grades 3-8; attendance for grades 6-12), treatment is defined by a school having an above-median pre-ban suspension rate relative to the estimation sample.

[27]These estimates can also be thought of as the difference in outcomes for SWDs in high-suspension schools relative to SWDs in low-suspension schools. As this paper is concerned with the differential effect of the ban, I focus on the comparison of groups within high-suspension schools.

[28]In contrast to the models presented in Section 5, I do not control for the pre-ban math and ELA scores given that these are my outcomes. As such, my sample is expanded to include those students that may not have pre-ban test scores. These samples (both test score and attendance rate samples) do not notably differ from each other and are available upon request.

[29]I include this control due to potential concerns regarding aSPtS.

[30]I conduct a robustness check to rule out this hypothesis in Section 8.3.

7.2 Mechanism: Effects of Ban-Induced Classifications on Student Outcomes

While the model outlined in section 7.1 allows for an understanding of differential changes for SWDs on average, it does not directly identify the impact of receiving a ban-induced classification. To identify the impacts of classifications that are plausibly ban-induced, I rely on an individual fixed effects intent-to-treat estimator as in Schwartz et al. (2021), and also account for the staggered nature of treatment—as suggested by O'Hagan and Stiefel (2024)—using the Callaway and Sant'Anna (2021) estimator given that not all students are classified with a disability in the same year. This resolves the negative weighting and "forbidden comparison" (Roth et al., 2023) problem that is likely present in prior and concurrent estimates of the impacts of disability classification derived from student fixed effect models (Hwang, 2024; Schwartz et al., 2021).

The Callaway and Sant'Anna (2021) estimator provides a variety of treatment effect aggregations. I rely on estimates of the group average treatment effect on the treated (GATT), that is, separate effect estimates for those classified in $2009, 2010, ..., 2015$. The GATT for each cohort is derived by using all valid 2×2 comparisons such that treated units are only ever compared to units that are never treated or not yet treated. For example, the effect of classification in 2012 for a student classified in 2010 would only be derived from those students who are never classified or classified in 2013, 2014, or 2015. For consistency, I estimate using never treated units (students always in general education) as my comparison.[31]

I estimate each GATT for each classification type (low-, stratified-, and high-status) and compare effects for the 2009, 2010, 2011, 2013, 2014, and 2015 cohorts relative to those classified in the last year before the ban (2012). Further, I estimate GATTs separately for those students classified in above- and below-median suspension schools, which allows me to parse out the contamination from aSPtS, which likely affected the efficacy of classification and, importantly, compare ban-induced classifications (in above-median suspension schools) to non-ban-induced classifications (in below-median suspension schools).[32] Ultimately, my parameter of interest is the difference in the effect of classification in a given year relative to the effect of classification in 2012 between students classified in an above-median suspension school and students classified in a below-median

[31]Estimates using not-yet-treated units (students not yet classified for special education) are uninformative due to power issues—for each cohort, there are minimal comparisons and, by construction, estimates cannot be derived for the 2015 cohort because every student is treated. As such, there would only be two post-period estimates.

[32]I expand on the assumptions of this strategy in Section 8.3.

suspension school. For each cohort c_j^{-2012}, classification status j, and classification setting s, where $s = \{\text{Above, Below}\}$, I generate:

$$\phi_{sc,j} = ATT(s,c)_j - ATT(s,2012)_j$$

and my ultimate parameter of interest for each classification cohort is:

$$\psi_{c,j} = \phi_{above,c,j} - \phi_{below,c,j}$$

which represents a cohort c_j's difference in the effect of being classified between high- and low-suspension schools, relative to that difference in 2012 for each classification status j.

8 Results: Effects of the Ban on Students with Disabilities

8.1 Differential Effects of the Suspension Ban on Students with Disabilities

I begin by estimating how certain disability statuses moderate the impact of the suspension ban on key outcomes. Figure 6 presents the coefficients on the interaction between a year, an indicator for being an above-median suspension school, and being a student with a specific disability classification. Panels A and B present changes in math and ELA test scores, respectively; Panel C presents coefficients for attendance rates. Tables 9, 10, and 11 present corresponding coefficients.

Generally, I reproduce results from Craig and Martin (2023), finding that the suspension ban improved outcomes on average—by 2015, GENs in above-median suspension schools experienced statistically significant 0.04sd increases in math scores. For SWDs on average, there are statistically insignificant decreases in test scores; however, for students with low-status disability classifications, ELA scores decreased by 0.25sd ($p < 0.05$) by 2015 and math scores decreased by 0.23sd ($p < 0.05$) in 2014 relative to their GEN peers.[33] For students with stratified and high-status classifications, there are insignificant differential changes in all outcomes. SWDs, regardless of classification, did not experience differential changes in attendance rates relative to GENs. Notably, I do find slight, yet statistically significant, decreases in attendance rates for GENs in above-median suspension schools—by 2015, attendance rates decreased by 0.57pp ($p < 0.01$) equivalent to just over one day in the academic year.[34]

[33]Estimates for 2013 and 2015 for math scores are insignificant.

[34]Because attendance is measured by the total number of days that a student attends out of all *possible* days, this means that a student that was no longer being suspended increases the denominator within the attendance rate calculation. This means that these drops in attendance could simply be mechanical assuming that a student attends the same number of days.

I also provide estimates for differential changes in outcomes for Black SWDs (Figure A7 and Table A11) and for Male SWDs (Figure A8 and Table A10). Black SWDs did not experience differential changes in ELA or math scores, but did experience a 0.67pp ($p < 0.05$) additional reduction in attendance rates relative to non-Black students, driven by high status classifications. Given that Black students were more likely to be transferred out of self-contained settings post-2012 (i.e., more affected by aSPtS), attributing these effects solely to the suspension ban is implausible. That Black students were disproprotionately removed from self-contained settings may suggest that Black students were disproportionately served in self-contained settings prior to aSPtS. Male SWDs experienced statistically significant differential reductions in ELA test scores by 0.08sd ($p < 0.001$) by 2015, driven by students with stratified status classifications. There was no statistically significant change in math test scores. By 2015, male students with stratified-status classifications experienced a 0.63pp ($p < 0.05$) reduction in attendance and those with high-status classifications reduced their attendance by 0.79pp ($p < 0.05$).

8.2 Effects of Ban-Induced Classifications

Figure 7 and Tables 12, 13, and 14 display estimates of the group average treatment effects on the treated for each cohort, separately for those students classified in an above- or a below-median suspension school for ELA, math, and attendance outcomes, respectively. As shown in Tables 12 and 13, there is minimal difference in the change in math and ELA scores between students who are classified in above- or below-median suspension school, regardless of cohort or classification type. However, there are differences in the effects of classification by classification status. For students with stratified-status classifications, test scores improve post-classification by 0.069sd or 0.093sd (both $p < 0.01$) for those in below- and above-median suspension schools, respectively. Students with high-status classifications also experienced improvements in ELA test scores after classification. However, for students with low-status classifications, ELA test scores decreased by 0.091sd ($p < 0.05$) in below-median suspension schools. A similar pattern emerges when looking at math test scores (Table 13). For students with low-status classifications, test scores tended to decrease post-classification, whereas for students with stratified status classifications, math test scores increased. For students with high-status classifications, there was no change in math test scores post-classification.

In contrast, attendance rates decreased upon classification, driven mostly by students receiving low-status classifications and by students in above-median suspension schools (Table 14). Importantly, these decreases are a continuation of a downward trend in attendance rates prior to classification (Figure A10). As such, these estimates should be interpreted with caution. The differences in ATTs between above- and below-median suspension schools are statistically significant for all classification types, with students classified in above-median suspension schools having greater decreases in attendance rates than those in below-median suspension schools. Students with low-status classifications experienced decreases in attendance by 6.6pp and 10.0pp (both $p < 0.001$) in below- and above-median suspension schools, respectively. Students with stratified status classifications experienced reductions in attendance by 2.5pp in below-median suspension schools and 3.6pp in above-median suspension schools (both $p < 0.001$). For students with high-status classifications, attendance decreased by 1.7pp ($p < 0.001$) in above-median suspension schools and did not change in below-median suspension schools. The differences in estimates between above- and below-median suspension schools are all statistically significant at at least $p < 0.05$.

These estimates provide a baseline for changes in outcomes post-classification, on average. To better define the impact of a ban-induced classification, I compare estimates of the ATT for each cohort (GATTs) relative to the effect of classification for those classified in 2012, then compare these differences to differences between students who are classified in above- and below-median suspension schools. Figure 8 depicts these estimates (i.e., ψ_c). Generally, there is minimal difference in the effect sizes between above- and below-median suspension schools relative to 2012; however, for students with stratified status classifications, the change in the effects of classification on math scores grew faster for students classified post-ban, relative to the change in effects for those students classified in below-median suspension schools. Put simply, this suggests that ban-induced classifications are generally not functioning significantly differently from non-ban-induced classifications and, if anything, may have provided services to students who would otherwise have been suspended as evidenced by slight improvements in the efficacy of stratified-status classifications in increasing math test scores.

8.3 Robustness to Identifying Assumptions

To plausibly derive the causal impact of the suspension ban on outcomes for SWDs relative to GENs, it must be the case that, absent the ban, the difference in outcomes between SWDs and GENs within above- and below-median suspension schools would have continued in parallel. I provide evidence that this assumption is plausible given the lack of pre-trends as depicted in Figure 6. I also show event study figures for the student fixed effects models (Figure A10) which suggest that there are limited pre-trends for test score results and that the pre-trends that do exist prior to classification are no different between those students classified in above- or below-median suspension schools. Further, Figure 8 demonstrates that there are no trends in the difference in estimates between above- and below-median suspension schools relative to the 2012 classification cohorts.

8.3.1 Addressing aSPtS and Ruling out Other Mechanisms

In Section 6.4.4, I showed that above- and below-median suspension schools were similarly impacted by the policy as measured by the change in the proportion of SWDs that were served in self-contained settings. It is plausible, however, that even if these schools were similarly impacted by the policy, their ability to respond to uniform increases in SWDs in inclusive settings may vary. Restricting my sample to SWDs only ever served in inclusive settings limits the bias of aSPtS' direct effect on students moving out of self-contained settings; however, it does not remove the potential peer effects of moving a previously self-contained student into an inclusive classroom when teachers' ability to handle these peer effects may vary systematically by school treatment status (i.e., above-median suspension schools were worse at providing inclusive services relative to below-median suspension schools). I show that this is likely not the case by estimating the difference in outcomes for students enrolled in schools more treated by aSPtS *and* more treated by the suspension ban.[35] These estimates are displayed in Table A12. For ELA scores, I find that my results remain robust—there is limited to no difference in ELA test score outcomes between those SWDs in high-suspension schools that have high or low pre-2012 rates of SWD exclusion. I find no

[35]This is a model that includes an interaction between year (2012 omitted), disability status, an indicator for a school with above/below-median pre-policy suspension rates, and an indicator for a school with above/below-median pre-policy proportion of students served in self-contained settings. This model is estimated on students who are GENs (reference group) or always served in inclusive settings.

significant difference in math scores or attendance rates for SWDs in any setting.

8.3.2 Composition Changes

Additionally, one may be concerned that the result of GENs' increasing ELA test scores with low-status SWDs' decreasing ELA test scores is a product of selection on the outcome. That is, one could imagine a world in which post-ban and in high-suspension schools, a teacher happens to classify students with the lowest test scores as a SWD with a low-status disability. As such, the result of lower test scores for low-status SWDs and higher test scores for GENs is manufactured based on selection based on the outcome. To address this concern, I demonstrate that there is minimal difference between pre-classification test scores for those students classified prior to the ban and those classified after the ban—any difference is not sufficient to completely account for the 0.25sd differential reduction in ELA test scores for students with low-status disabilities. If anything, students classified post-ban have slightly higher math test scores than those classified pre-ban. Table A13 presents these results.

8.3.3 Controls with Student Fixed Effects

I also examine whether results from student fixed effect models change when including students who are classified in a self-contained setting. I expand my estimation sample to include any SWD (regardless of service setting) and re-estimate the student fixed effect model with the Callaway and Sant'Anna (2021) estimator and find that my estimates of the difference in GATTs between above- and below-median suspension schools relative to the 2012 difference do not notably change (see Figure A9). As such, there is minimal difference in the estimates between how students always classified in inclusive settings fare in above- or below-median suspension schools relative to those students who may sometimes be in self-contained settings.

9 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. 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, estimating whether the ban induced students into more restrictive environments (i.e., as a true method of exclusion) 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. I use sample restrictions to account for the direct impact of this policy on SWDs moving out of self-contained settings and conduct heterogeneity tests to address the potential presence of peer effects of SWD inclusion. While this may be a limitation, it also highlights the importance of caution when constructing research designs in complex policy environments—a policy that may be seemingly unrelated (i.e., aSPtS) should be explicitly and carefully considered.

Lastly, estimates regarding the impact of a ban-induced classification are subject to issues of power, limiting precise estimates. Further, I am, as with any research within the realm of special education and suspension reform, unable to observe the true underlying behavior or disability of a students. As such, I provide suggestive evidence throughout Section 8 as to how SWDs on average and with potentially ban-induced classifications fare in the wake of the suspension ban. Similarly, defining what constitutes a "ban-induced" classification relies on the assumption that those classifications that occurred in above-median suspension schools are all, indeed, a result of the ban. This assumption seems unlikely, but, absent observing the "true," underlying disability, it is unable to be tested.

10 Discussion

Understanding how discipline bans function—and their potential pitfalls—is of utmost importance to ensure that 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. Ideally, these policies would indeed be costless, relying only on a change in policy code to function. However, the results presented in this paper suggest an unexamined cost of suspension ban: classification of students with more stigmatizing disabilities and, further, reductions in test scores for students with these disabilities on average. I find that the New York City suspension ban induced an increase in higher stigma special education classifications, which was concentrated among male students. Further, the ban also increased gender and racial disproportionality in classification on average. Insofar as special education services remain costly (and resources spread thin), a suspension ban that induces increases in classifications acts as a hidden cost to these policies.

I show that students with low-status classifications experienced on average decreases in math and ELA test scores, whereas students with stratified and high-status classifications experienced no differential change relative to their GEN peers. I rule out other mechanisms and policies that could potentially be driving this result, suggesting that the negative changes in low-status SWDs' test scores are largely due to a differential response to ban-induced environmental changes. The reduction in SWDs' achievement also represents a hidden cost—SWD test score reductions by 0.25sd is equivalent to cutting school spending by nearly $6500 per pupil.[36].

Simultaneously, I also show that those classifications that are plausibly ban-induced do not function differently than non-ban-induced classifications, suggesting that suspension may have prohibited these students from receiving services. Taken together, I provide evidence that is congruent with two seemingly opposing literatures: one that suggests special education functions as a new form of exclusion (Blanchett, 2006, 2010; Kim et al., 2010) and another that documents benefits of special education (O'Hagan and Stiefel, 2024; Ballis and Heath, 2021b; Schwartz et al., 2021; Elder et al., 2021). In this paper, I demonstrate that even if special education may be used as a mechanism of exclusion (i.e., ban-induced classifications), these classifications are likely preferable to

[36]This is calculated using Jackson et al. (2021)'s finding of $1000 cuts in per-pupil spending leading to $\sim$0.039sd decrease on average in test scores. As such, a reduction of 0.25sd is equivalent to $(\frac{0.25}{0.039}) \times \$1000$.

the counterfactual world in which a student is instead suspended. Although, classification is likely not an efficient strategy insofar other methods (e.g., RPs, progressive discipline, etc.) utilize less resources and benefit students.

I also build upon the literature that suggests that suspension bans, when implemented with fidelity, improve outcomes on average (Craig and Martin, 2023; Cleveland, 2023; Steinberg and Lacoe, 2018a; Karger and Komisarow, 2024), but contribute new evidence as to how vulnerable groups may be negatively impacted. 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 strategy to address behavior (suspension) results in a substitution to another (special education). I also contribute to the relatively nascent quantitative literature on the effects of special education and provide the first estimates using an estimator that accounts for the staggered treatment. I show that post-classification, ELA and math scores tend to increase.

These results are not to say that the suspension ban was ineffective, but instead they point to where future policy may wish to explicitly consider how students with disabilities may be impacted and prepare to allocate additional resources to special education programming. Future policy should explicitly consider the potential avenues that disciplinarians may take to address 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; however, this came at the cost of worsened test score outcomes for SWDs with low-status disabilities. 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 and immediately trained in alternative techniques. 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.

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

Baker, A. (2012, August). New Code Aims to Ease Suspensions of Students. *The New York Times*.

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.

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of econometrics 225*(2), 200–230.

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.

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.

Fronius, T., H. Persson, S. Guckenburg, N. Hurley, and A. Petrosino (2016). Restorative justice in us schools: A research review. *WestEd*.

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. (2024, April). Label to help: The effects of special education on student outcomes.

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.

Jackson, C. K., C. Wigger, and H. Xiong (2021, May). Do school spending cuts matter? evidence from the great recession. *American Economic Journal: Economic Policy 13*(2), 304–35.

Karger, E. and S. Komisarow (2024). *Ending Early Grade Suspensions*. Ph. D. thesis, Duke University.

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.

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

Lincove, J. A., C. Mata, and K. E. Cortes (2024). The effects of a statewide ban on school suspensions. edworkingpaper no. 24-1004. *Annenberg Institute for School Reform at Brown University*.

Maciver, D., M. Rutherford, S. Arakelyan, J. M. Kramer, J. Richmond, L. Todorova, D. Romero-Ayuso, H. Nakamura-Thomas, M. Ten Velden, I. Finlayson, et al. (2019). Participation of children with disabilities in school: A realist systematic review of psychosocial and environmental factors. *PloS one 14*(1), e0210511.

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.

O'Hagan, K. G. and L. Stiefel (2024). Does special education work? a systematic literature review of evidence from administrative data. *Remedial and Special Education*, 07419325241244485.

O'Hagan, K. G., L. Stiefel, and A. E. Schwartz (2023). The academic effects of moving to middle school on students with disabilities relative to their general education peers. *Educational Evaluation and Policy Analysis*, 01623737241229990.

Pope, N. G. and G. W. Zuo (2023). Suspending suspensions: The education production consequences of school suspension policies. *The Economic Journal 133*(653), 2025–2054.

Rodriguez, L. A. and R. O. Welsh (2022). The dimensions of school discipline: Toward a comprehensive framework for measuring discipline patterns and outcomes in schools. *AERA Open 8*, 23328584221083669.

Roth, J., P. H. Sant'Anna, A. Bilinski, and J. Poe (2023). What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics 235*(2), 2218–2244.

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.

Simpson, R. L., R. L. Peterson, and C. R. Smith (2011). Critical educational program components for students with emotional and behavioral disorders: Science, policy, and practice. *Remedial and special education 32*(3), 230–242.

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 (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.

Sun, M., C. A. Candelaria, D. Knight, Z. LeClair, S. E. Kabourek, and K. Chang (2024). The effects and local implementation of school finance reforms on teacher salary, hiring, and turnover. *Educational Evaluation and Policy Analysis*, 01623737231213880.

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.

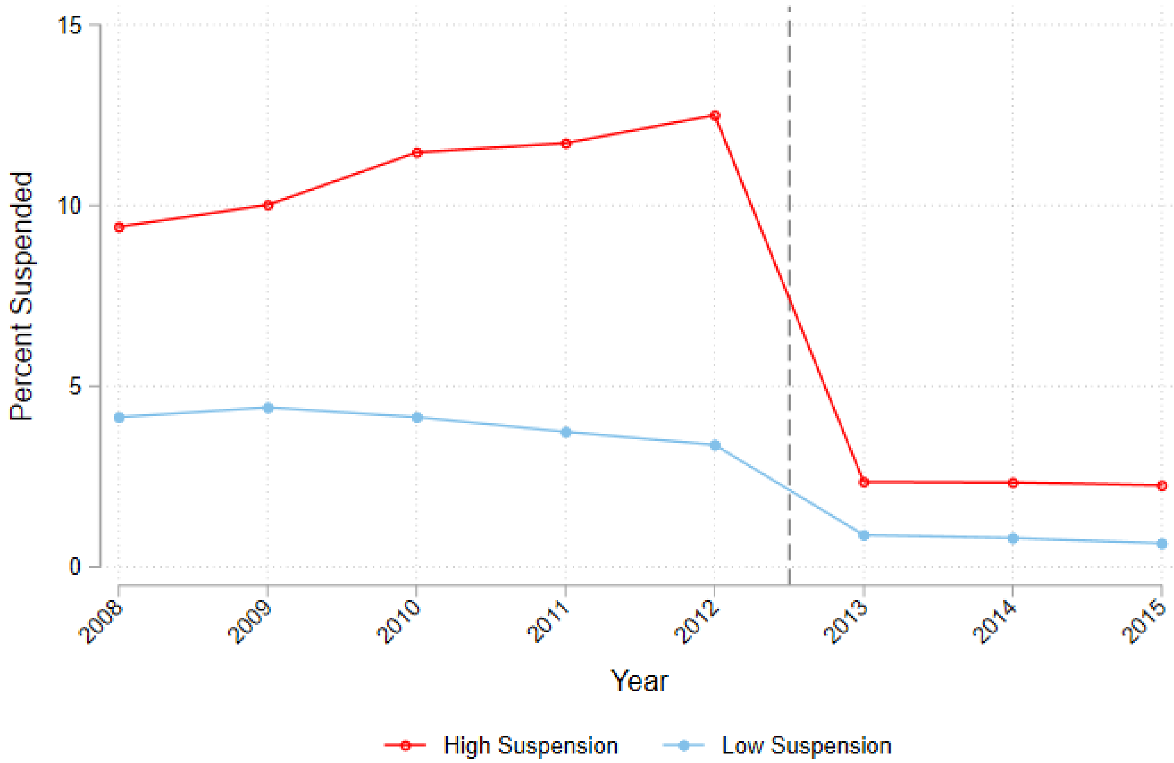
USDOE (2018). Individuals with disabilities education act; regulations part b, subpart a, sec 300.8 (c). https://sites.ed.gov/idea/regs/b/a/300.8/c.

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.

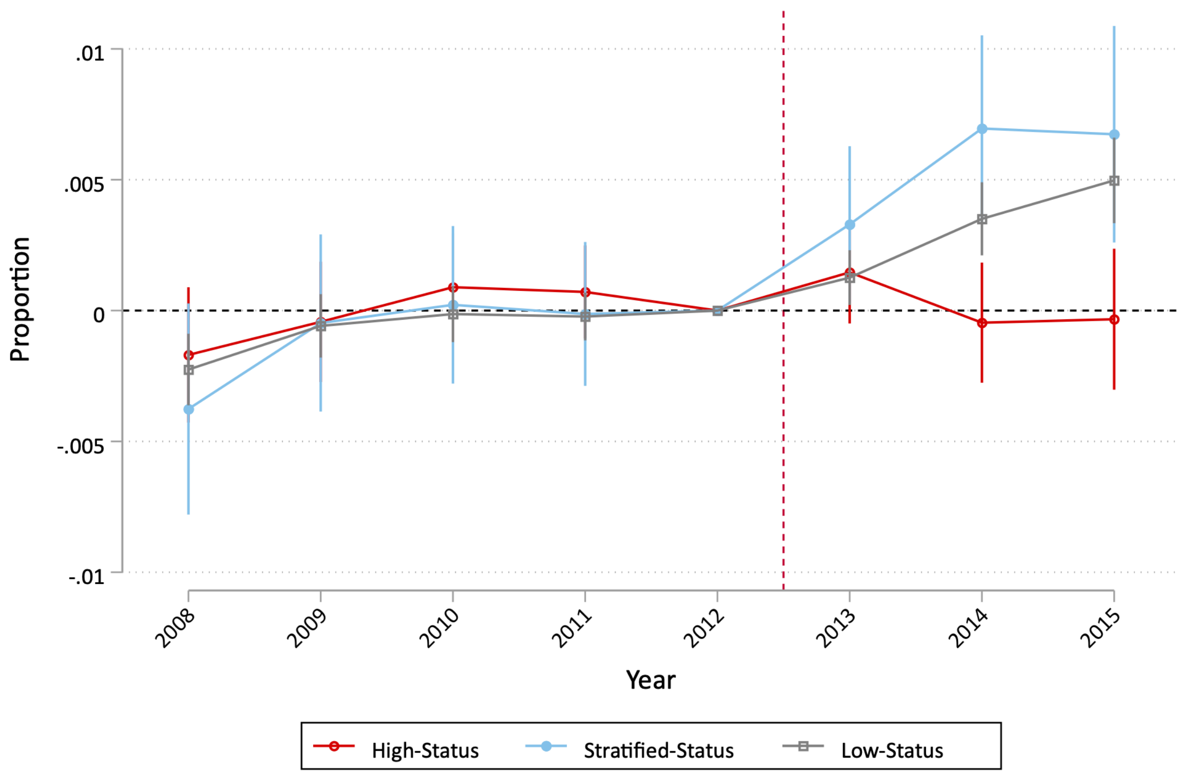
Welsh, R. O. and S. Little (2018b). The school discipline dilemma: A comprehensive review of disparities and alternative approaches. *American Educational Research Association 88*(5), 752–794.

Figure 1: Identifying Variation: Percent Suspended by Treatment Status



Notes. This figure depicts the raw school-level suspension rates for those schools that have above-median suspension rates and those schools that have below-median suspension rates.

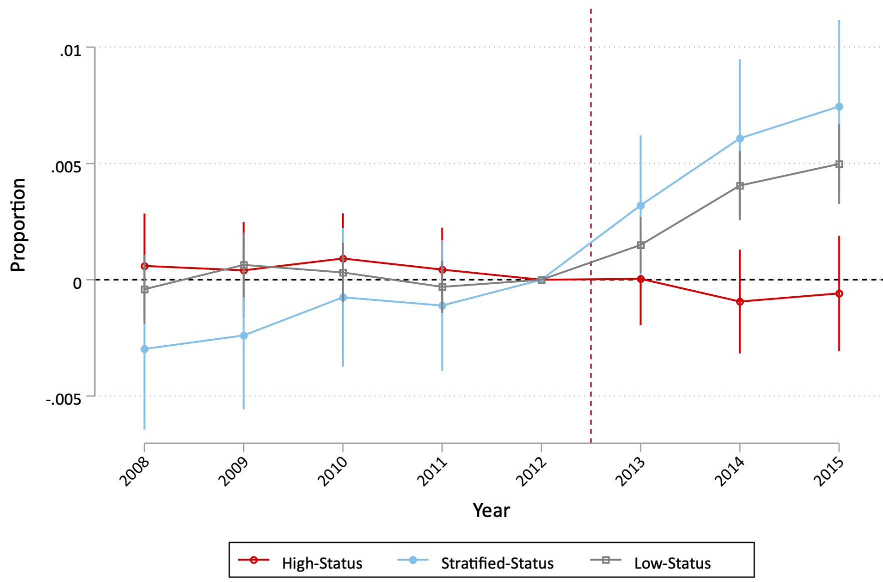
Figure 2: Event Study Difference-in-Difference Estimates: Above−Below Median Suspension



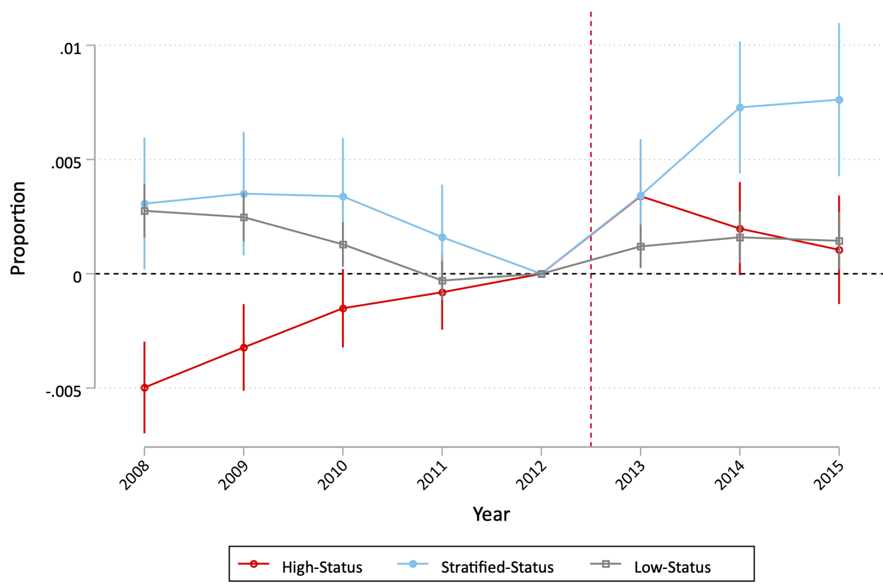
Notes. Estimates presented correspond to those in Table 3, respectively. Estimates indicate the average change in classification rates relative to the baseline academic year (2011-12) for all middle and high school students. 95% confidence intervals are provided. Panel A shows estimates from interrupted time series and Panel B shows estimates from the event study dosage difference-in-differences model.

Figure 3: Treatment by Race and Sex

(a) Black—Non-Black Students



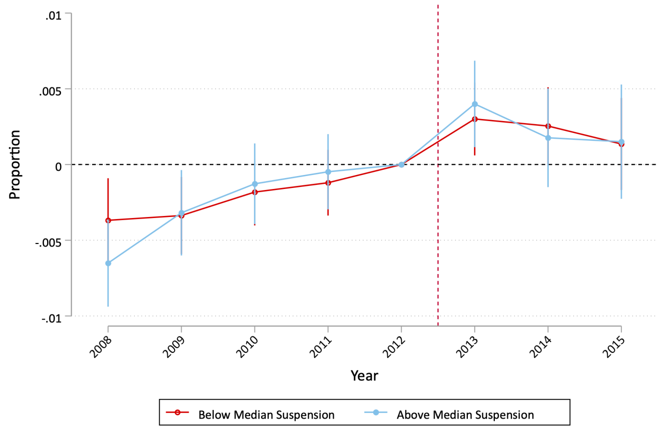
(b) Male—Female Students



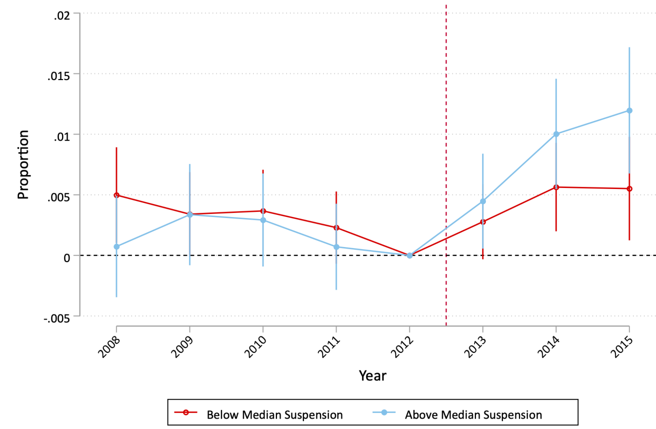
Notes. Estimates presented correspond to the trend over time disproportionality estimates in Table 4 Panels A and B, respectively. Estimates indicate the average change in classification rates relative to the baseline academic year (2011-12) for all middle and high school students. Panel A compares male students to female students relative to the difference in 2012 and Panel B compares Black students to non-Black students relative to the difference in 2012. 95% confidence intervals are provided.

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

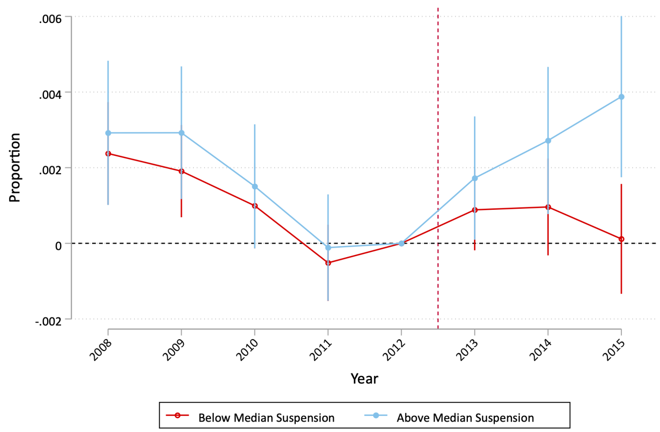
(a) High-Status



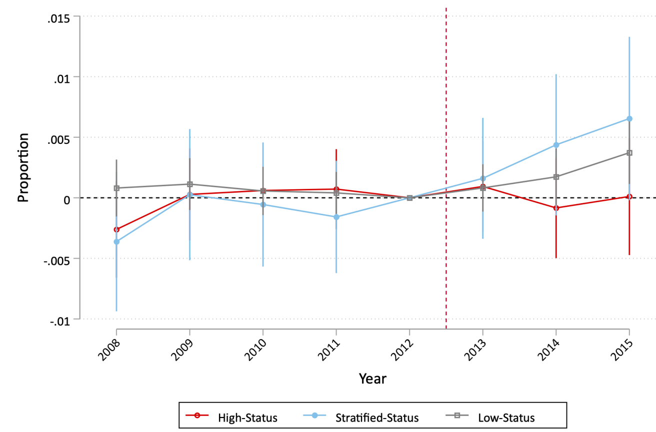
(b) Stratified Status



(c) Low-Status



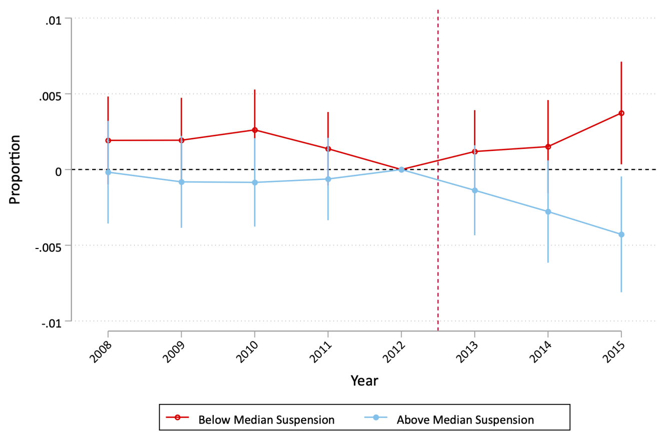
(d) Triple Difference



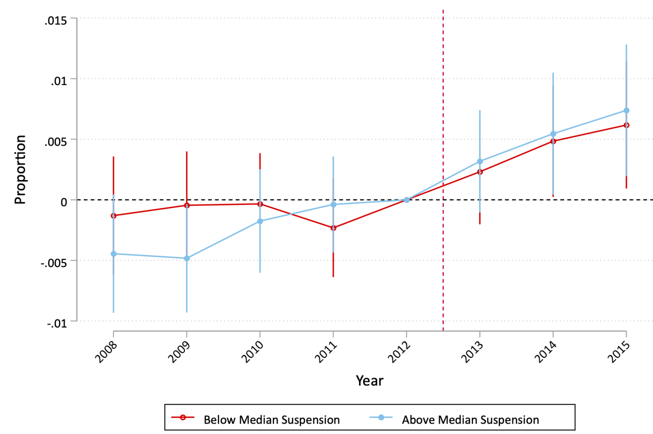
Notes. Estimates in Panels A-C presented represent the male-female gap in classification within a high or low-suspension school (above or below the 2011-12 district median suspension rate of 6 percent). Estimates in Panel D indicate the triple difference coefficient. 95% confidence intervals are displayed.

Figure 5: Black−Non-Black Students in High and Low Suspension Schools

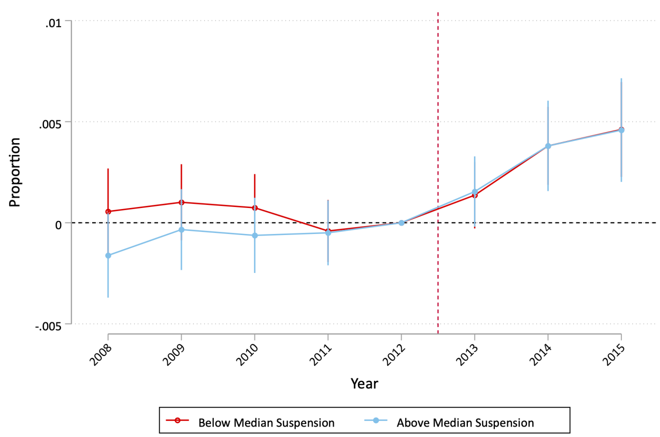
(a) High-Status



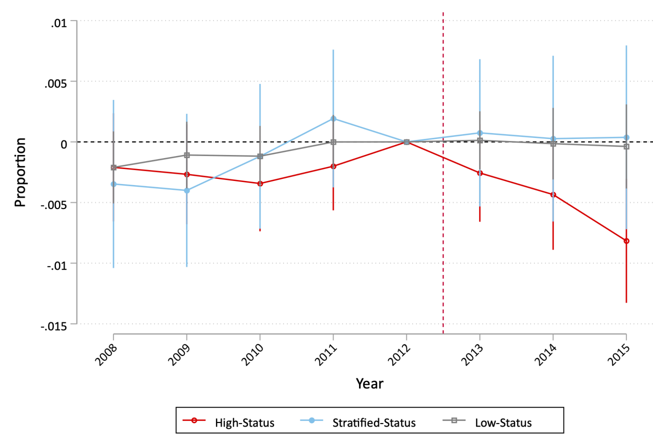
(b) Stratified Status



(c) Low-Status



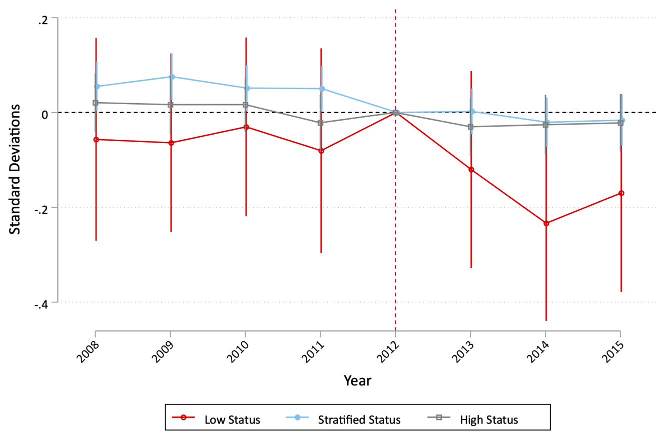
(d) Triple Difference



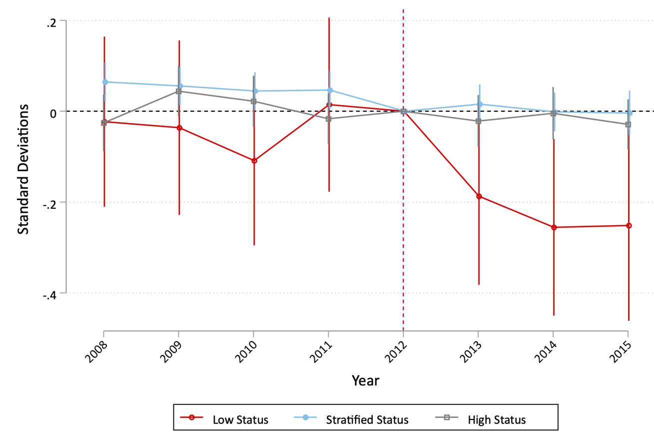
Notes. Estimates in Panels A-C presented represent the Black-non-Black gap in classification within a high or low-suspension school (above or below the 2011-12 district median suspension rate of 6 percent). Estimates in Panel D indicate the triple difference coefficient. 95% confidence intervals are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

Figure 6: Differential Effects of the Suspension Ban on SWDs

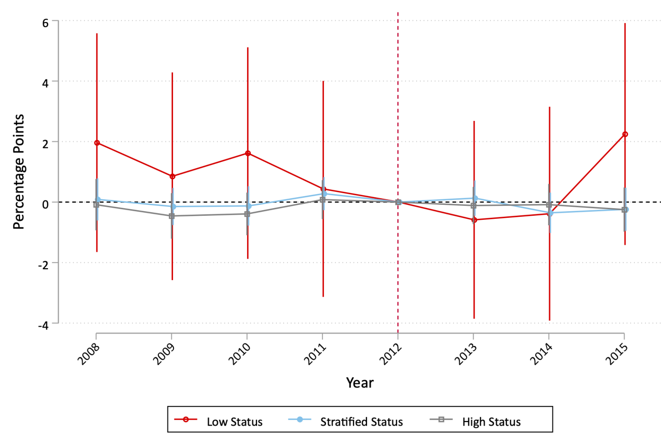
(a) Math Scores



(b) ELA Scores

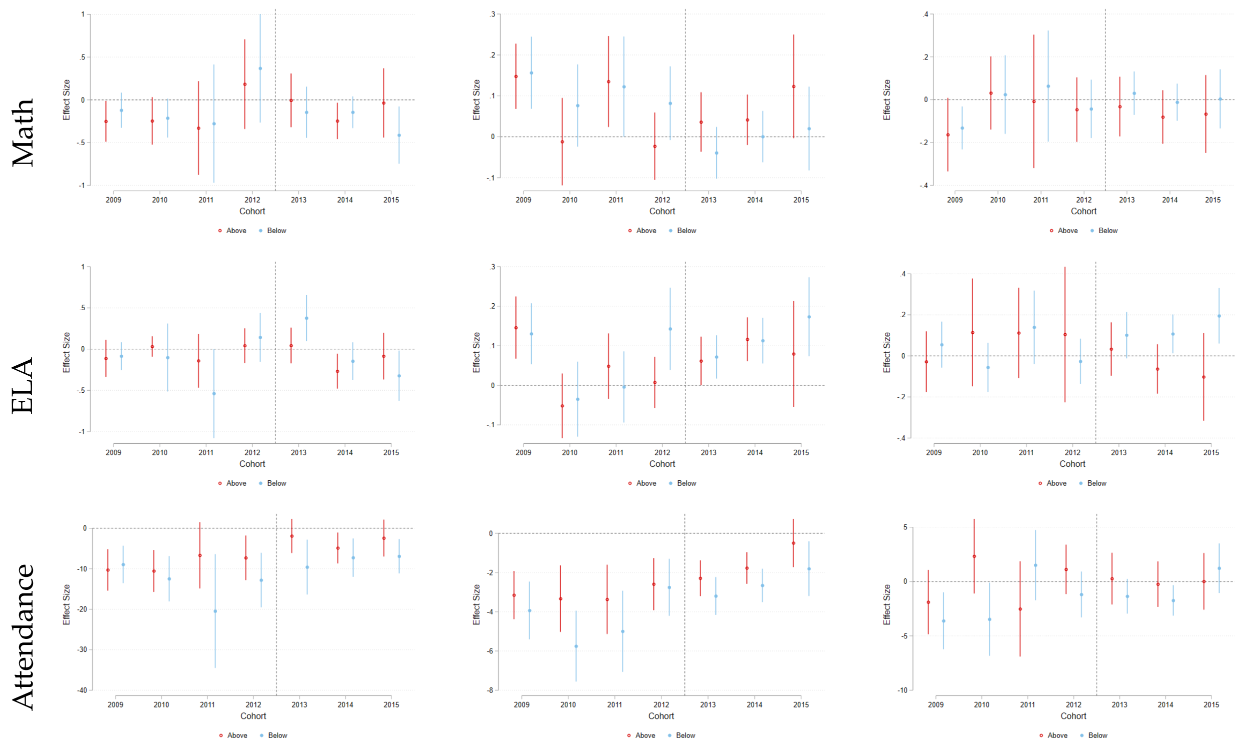


(c) Attendance



Notes. Estimates reflect the coefficients on an interaction between a year indicator, a classification indicator, and an indicator for being an above-median suspension school relative to 2012. Math and ELA scores are standardized to mean 0 and standard deviation of 1. Math and ELA estimates are derived from students in grades 6-8; attendance from students in grades 6-12. All samples are restricted to SWDs that are only ever served in inclusive settings. The vertical dashed red line indicates the year of the policy change (2012). 95% confidence intervals are displayed.

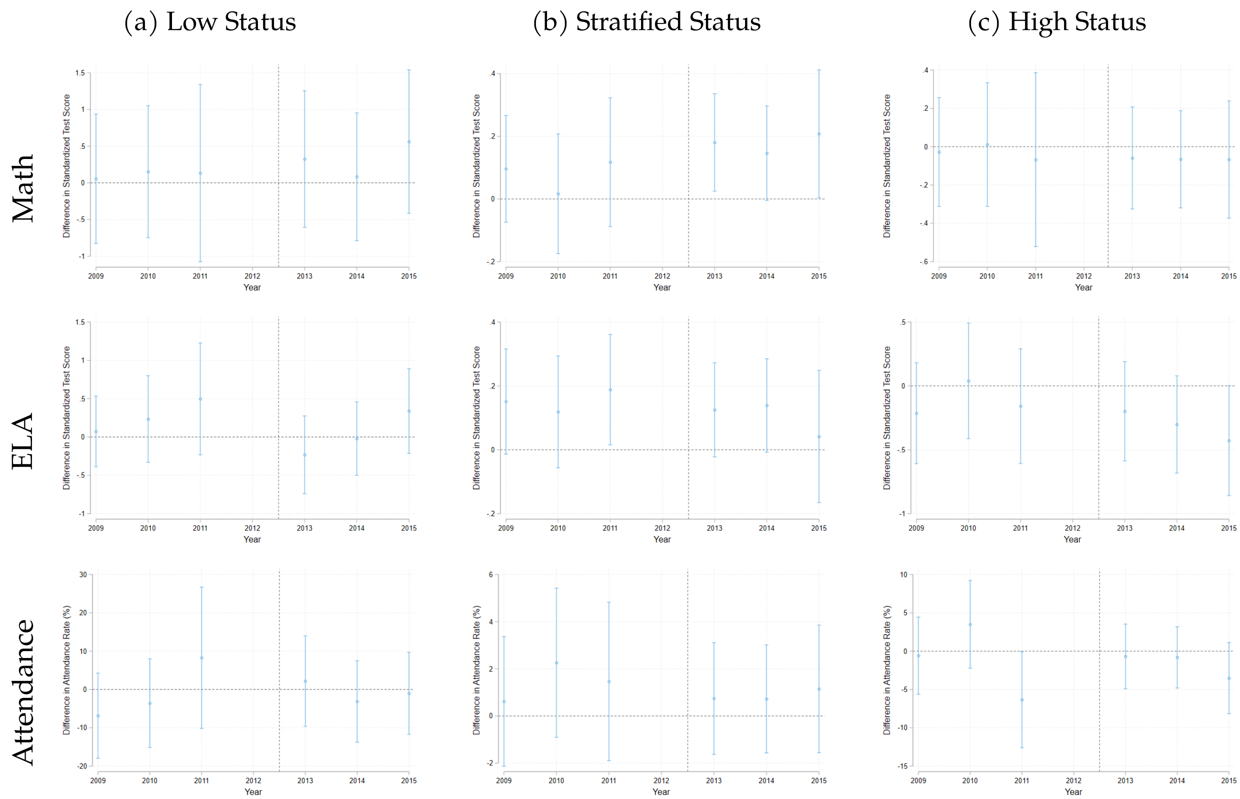
Figure 7: Group Average Treatment Effects on the Treated



(a) Low Status (b) Stratified Status (c) High Status

Notes. Estimates reflect the effect of receiving a subjective disability classification in a specific cohort as derived by an intent-to-treat student fixed effects model using the Callaway and Sant'Anna (2021) estimator. Separate estimates are provided for those classified in above-median suspension schools and for those classified in below-median suspension schools. The comparison group in all estimates are students who never receive a disability classification. All samples are restricted to SWDs that are only ever served in inclusive settings. The vertical dashed line indicates the year of the policy change (2012). 95% confidence intervals are displayed.

Figure 8: Difference in Group Average Treatment Effects on the Treated between Above- and Below-Median Suspension Schools Relative to 2012



Notes. Estimates reflect the difference in the effect of receiving a subjective disability classification in a specific cohort as derived by an intent-to-treat student fixed effects model using the Callaway and Sant'Anna (2021) estimator relative to the effect size for those students classified in 2012. The comparison group in all estimates are students who never receive a disability classification. All samples are restricted to SWDs that are only ever served in inclusive settings. All models include controls for whether a student was held back and for a student's grade. The vertical dashed line indicates the year of the policy change (2012). 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	0.047	0.041
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	1,214,143	2,588,870
Schools	312	410	722

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***	-0.001
	(0.001)	(0.002)	(0.001)
2009	0.001***	0.011***	-0.001
	(0.000)	(0.001)	(0.001)
2010	0.002***	0.010***	0.001
	(0.000)	(0.001)	(0.001)
2011	0.000	0.002**	-0.000
	(0.000)	(0.001)	(0.000)
2013	0.001***	0.009***	0.007***
	(0.000)	(0.001)	(0.001)
2014	0.003***	0.016***	0.008***
	(0.000)	(0.001)	(0.001)
2015	0.004***	0.018***	0.010***
	(0.000)	(0.001)	(0.001)
Observations	2,588,870	2,588,870	2,588,870
Clusters	2,899	2,899	2,899

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 NYCPS, 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: Above − Below Median Suspension

	Low-Status	**Stratified Status**	**High-Status**
2012 Baseline Difference	*0.007*	*0.031*	*0.012*
2013	0.001**	0.008***	0.007***
	(0.000)	(0.001)	(0.001)
2014	0.002***	0.013***	0.008***
	(0.000)	(0.001)	(0.001)
2015	0.002***	0.016***	0.010***
	(0.000)	(0.001)	(0.001)
Above Median $\times$2013	0.001*	0.003*	0.001
	(0.001)	(0.002)	(0.001)
Above Median $\times$2014	0.004***	0.007***	-0.000
	(0.001)	(0.002)	(0.001)
Above Median $\times$2015	0.005***	0.007**	-0.000
	(0.001)	(0.002)	(0.001)
Observations	2,588,870	2,588,870	2,588,870
Clusters	2,899	2,899	2,899

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 NYCPS, mean pre-ban ELA and math test scores, 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: Heterogeneity by Race and Sex

	Low-Status	Stratified Status	High-Status
Panel A. Black-Non-Black			
2012 Baseline Difference	*0.011*	*0.023*	*-0.003*
2013	0.001*	0.008***	0.007***
	(0.000)	(0.001)	(0.001)
2014	0.002***	0.013***	0.008***
	(0.000)	(0.001)	(0.001)
2015	0.001***	0.015***	0.010***
	(0.000)	(0.001)	(0.001)
Black $\times$2013	0.001*	0.003*	0.000
	(0.001)	(0.002)	(0.001)
Black $\times$2014	0.004***	0.006***	-0.001
	(0.001)	(0.002)	(0.001)
Black $\times$2015	0.005***	0.007***	-0.001
	(0.001)	(0.002)	(0.001)
Panel B. Male-Female			
2012 Baseline Difference	*0.012*	*0.036*	*0.032*
2013	0.000	0.007***	0.005***
	(0.000)	(0.001)	(0.001)
2014	0.002***	0.011***	0.006***
	(0.000)	(0.001)	(0.001)
2015	0.002***	0.013***	0.009***
	(0.000)	(0.001)	(0.001)
Male $\times$2013	0.001*	0.003**	0.003***
	(0.000)	(0.001)	(0.001)
Male $\times$2014	0.002**	0.007***	0.002
	(0.001)	(0.001)	(0.001)
Male $\times$2015	0.001*	0.008***	0.001
	(0.001)	(0.002)	(0.001)
Observations	2,588,870	2,588,870	2,588,870
Clusters	2,899	2,899	2,899

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 NYCPS, 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.

Table 5: Triple Difference Estimates: Male Students or Black Students in Above Median Suspension Schools

	Low Status	**Stratified Status**	**High Status**
Panel A. Black-Non-Black			
2013 × Above Median × Black	0.000	0.001	-0.003
	(0.001)	(0.003)	(0.002)
2014 × Above Median × Black	-0.000	0.000	-0.004
	(0.002)	(0.003)	(0.002)
2015 × Above Median × Black	-0.000	0.000	-0.008**
	(0.002)	(0.004)	(0.003)
Panel B. 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	2,588,870	2,588,870	2,588,870
Clusters	2,899	2,899	2,899

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 NYCPS, 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%. Estimates reflect the coefficient on an interaction between a year indicator, an indicator for attending an above-median suspension school, and an indicator for being Male (Black). All estimates are relative to 2012.

Table 6: Falsification Test: Impacts of the Ban on Hard Classifications

	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	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Group × 2013		0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)
Group × 2014		0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)
Group × 2015		-0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)
Observations	2,588,870	2,588,870	2,588,870	2,588,870
Clusters	2,899	2,899	2,899	2,899

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 NYCPS, 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%. All estimates are relative to 2012. Hard classifications include deaf-blindness, hearing impairment, visual impairment, orthopedic impairment, traumatic brain injury.

Table 7: Changes in SWDs Served in Self-Contained Settings

	Above/Below	Male/Female	Black/Non-Black
2013	-0.066***	-0.060***	-0.061***
	(0.006)	(0.005)	(0.005)
2014	-0.061***	-0.059***	-0.055***
	(0.007)	(0.006)	(0.005)
2015	-0.065***	-0.065***	-0.060***
	(0.008)	(0.007)	(0.007)
Group × 2013	0.004	-0.008	-0.012*
	(0.008)	(0.005)	(0.005)
Group × 2014	0.008	-0.002	-0.016**
	(0.008)	(0.005)	(0.006)
Group × 2015	0.006	-0.003	-0.023***
	(0.008)	(0.005)	(0.006)
Observations	521,697	521,697	521,697

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. The sample is restricted to any student with a disability, with the outcome variable being an indicator for receiving services in a self-contained setting. Above-median indicates that a school's suspension rate in the 2011-12 AY was greater than the district median of 6%. All estimates are relative to 2012.

Table 8: Changes in SWDs Served in Self-Contained Settings: Triple Difference

	Male/Female	Black/Non-Black
2013	-0.064***	-0.061***
	(0.007)	(0.006)
2014	-0.064***	-0.055***
	(0.008)	(0.007)
2015	-0.064***	-0.059***
	(0.009)	(0.008)
Demographic × 2013	-0.002	-0.018*
	(0.007)	(0.008)
Demographic × 2014	0.004	-0.024**
	(0.007)	(0.009)
Demographic × 2015	-0.002	-0.023*
	(0.007)	(0.009)
Above × Demographic × 2013	-0.011	0.009
	(0.009)	(0.011)
Above × Demographic × 2014	-0.010	0.012
	(0.009)	(0.011)
Above × Demographic × 2015	-0.002	-0.000
	(0.010)	(0.012)
Observations	521,697	521,697

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. The sample is restricted to any student with a disability, with the outcome variable being an indicator for receiving services in a self-contained setting. Above-median indicates that a school's suspension rate in the 2011-12 AY was greater than the district median of 6%. All estimates are relative to 2012.

Table 9: Differential Effects of the Suspension Ban on SWDs: Math Scores

	Any	Low	Stratified	High
2013	0.011	0.010	0.011	0.009
	(0.009)	(0.009)	(0.009)	(0.009)
2014	0.036*	0.036*	0.036*	0.034*
	(0.015)	(0.015)	(0.015)	(0.015)
2015	0.053*	0.047*	0.048*	0.045*
	(0.021)	(0.022)	(0.021)	(0.021)
Above × 2013	-0.001	-0.001	-0.001	-0.001
	(0.013)	(0.013)	(0.013)	(0.013)
Above × 2014	0.010	0.009	0.009	0.009
	(0.018)	(0.018)	(0.018)	(0.018)
Above × 2015	0.038*	0.036*	0.037*	0.036*
	(0.018)	(0.018)	(0.018)	(0.018)
Disability × Above × 2013	-0.015	-0.120	0.002	-0.030
	(0.020)	(0.106)	(0.025)	(0.031)
Disability × Above × 2014	-0.034	-0.233*	-0.021	-0.026
	(0.023)	(0.105)	(0.027)	(0.032)
Disability × Above × 2015	-0.026	-0.170	-0.016	-0.022
	(0.022)	(0.106)	(0.027)	(0.031)
Observations	1,227,979	1,100,023	1,178,496	1,142,038

Notes. Standard errors in parentheses ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$) and are clustered at the school-grade level. Estimates reflect the coefficients on an interaction between a year indicator, a classification indicator, and an indicator for being an above-median suspension school relative to 2012. All models use school-by-grade fixed effects and control for student gender, race, ELL status, FRPL eligibility, and years present in NYCPS. Math scores are standardized to a mean of 0 and standard deviation of 1. All samples are restricted to SWDs that are only ever served in inclusive settings and to students with available test scores (grades 6-8). students who are not classified with any disability are the reference group.

Table 10: Differential Effects of the Suspension Ban on SWDs: ELA Scores

	Any	Low	Stratified	High
2013	0.010	0.008	0.010	0.008
	(0.007)	(0.007)	(0.007)	(0.007)
2014	0.016	0.012	0.014	0.011
	(0.010)	(0.010)	(0.010)	(0.010)
2015	0.010	0.000	0.005	-0.000
	(0.014)	(0.014)	(0.014)	(0.014)
Above × 2013	-0.022*	-0.022*	-0.022*	-0.022*
	(0.010)	(0.010)	(0.010)	(0.010)
Above × 2014	-0.010	-0.010	-0.010	-0.010
	(0.012)	(0.012)	(0.012)	(0.012)
Above × 2015	0.002	0.002	0.002	0.001
	(0.013)	(0.013)	(0.013)	(0.013)
Disability × Above × 2013	-0.007	-0.187	0.016	-0.022
	(0.019)	(0.099)	(0.022)	(0.029)
Disability × Above × 2014	-0.016	-0.256*	-0.001	-0.005
	(0.019)	(0.099)	(0.022)	(0.029)
Disability × Above × 2015	-0.018	-0.252*	-0.004	-0.029
	(0.020)	(0.107)	(0.025)	(0.028)
Observations	1,222,612	1,094,118	1,172,845	1,136,349

Notes. Standard errors in parentheses ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$) and are clustered at the school-grade level. Estimates reflect the coefficients on an interaction between a year indicator, a classification indicator, and an indicator for being an above-median suspension school relative to 2012. All models use school-by-grade fixed effects and control for student gender, race, ELL status, FRPL eligibility, and years present in NYCPS. ELA scores are standardized to a mean of 0 and standard deviation of 1. All samples are restricted to SWDs that are only ever served in inclusive settings and to students with available test scores (grades 6-8). students who are not classified with any disability are the reference group.

Table 11: Differential Effects of the Suspension Ban on SWDs: Attendance

	Any	Low	Stratified	High
2013	0.202*	0.209*	0.201*	0.202*
	(0.090)	(0.090)	(0.090)	(0.089)
2014	0.248*	0.265*	0.249*	0.259*
	(0.125)	(0.125)	(0.125)	(0.125)
2015	1.094***	1.134***	1.097***	1.122***
	(0.149)	(0.150)	(0.149)	(0.149)
Above × 2013	-0.327*	-0.324*	-0.328*	-0.322*
	(0.131)	(0.130)	(0.131)	(0.130)
Above × 2014	-0.747***	-0.727***	-0.740***	-0.721***
	(0.172)	(0.171)	(0.171)	(0.170)
Above × 2015	-0.570**	-0.529**	-0.551**	-0.524**
	(0.182)	(0.182)	(0.182)	(0.181)
Disability × Above × 2013	-0.009	-0.586	0.130	-0.113
	(0.237)	(1.667)	(0.297)	(0.314)
Disability × Above × 2014	-0.318	-0.384	-0.354	-0.084
	(0.268)	(1.802)	(0.340)	(0.349)
Disability × Above × 2015	-0.082	2.249	-0.239	-0.248
	(0.280)	(1.871)	(0.362)	(0.372)
Observations	3,415,156	3,128,624	3,320,625	3,201,673

Notes. Standard errors in parentheses ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$) and are clustered at the school-grade level. Estimates reflect the coefficients on an interaction between a year indicator, a classification indicator, and an indicator for being an above-median suspension school relative to 2012. All models use school-by-grade fixed effects and control for student gender, race, ELL status, FRPL eligibility, and years present in NYCPS. Attendance rates are the percent of days attended out of allowable/possible days and are available for all students. All samples are restricted to SWDs that are only ever served in inclusive settings. students who are not classified with any disability are the reference group.

Table 12: Group Average Treatment Effects of Classification: ELA

	Low-Status			**Stratified Status**			**High Status**		
Classification Cohort	*Below*	*Above*	*Difference*	*Below*	*Above*	*Difference*	*Below*	*Above*	*Difference*
2009	-0.085	-0.113	-0.028	0.130***	0.146***	0.016	0.055	-0.028	-0.083
	(0.086)	(0.115)	(0.143)	(0.039)	(0.040)	(0.056)	(0.057)	(0.075)	(0.095)
2010	-0.102	0.032	0.134	-0.035	-0.052	-0.017	-0.056	0.114	0.170
	(0.210)	(0.064)	(0.220)	(0.048)	(0.042)	(0.064)	(0.061)	(0.134)	(0.147)
2011	-0.538	-0.141	0.396	-0.004	0.049	0.053	0.140	0.112	-0.028
	(0.276)	(0.167)	(0.323)	(0.046)	(0.042)	(0.062)	(0.091)	(0.112)	(0.144)
2012	0.143	0.042	-0.101	0.143**	0.008	-0.135*	-0.026	0.105	0.131
	(0.152)	(0.107)	(0.186)	(0.053)	(0.033)	(0.062)	(0.056)	(0.168)	(0.178)
2013	0.377**	0.044	-0.333	0.072*	0.062*	-0.010	0.101	0.034	-0.068
	(0.142)	(0.111)	(0.180)	(0.028)	(0.031)	(0.042)	(0.058)	(0.066)	(0.088)
2014	-0.146	-0.268*	-0.122	0.113***	0.116***	0.003	0.107*	-0.063	-0.171*
	(0.116)	(0.108)	(0.159)	(0.029)	(0.028)	(0.041)	(0.048)	(0.061)	(0.078)
2015	-0.323*	-0.085	0.238	0.173***	0.079	-0.094	0.195**	-0.102	-0.297*
	(0.155)	(0.145)	(0.212)	(0.051)	(0.068)	(0.085)	(0.069)	(0.109)	(0.129)
Simple ATTs	-0.091*	-0.037	-0.054	0.069**	0.093**	-0.023	0.014	0.077**	-0.063
	(0.045)	(0.060)	(0.076)	(0.014)	(0.015)	(0.021)	(0.042)	(0.024)	(0.048)

Notes. Standard errors in parentheses ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$). Estimates reflect estimates the group average treatment effect on the treated for each classification cohort as derived from a intent-to-treat student fixed effect model using the Callaway and Sant'Anna (2021) estimator. Estimates for students classified in an above-median suspension school and for those classified in a below-median suspension school, as well as the difference in effect sizes are displayed. Math and ELA scores are standardized to mean 0 and standard deviation of 1 and are available for students in grades 6-8. Attendance rates are the percent of days attended out of allowable/possible days and are available for all students. Students who are never classified with a disability are the omitted group.

Table 13: Group Average Treatment Effects of Classification: Math

Classification Cohort	**Low-Status**			**Stratified Status**			**High Status**		
	Below	*Above*	*Difference*	*Below*	*Above*	*Difference*	*Below*	*Above*	*Difference*
2009	-0.122	-0.251*	-0.130	0.156***	0.148***	-0.009	-0.132**	-0.163	-0.031
	(0.105)	(0.122)	(0.161)	(0.045)	(0.041)	(0.061)	(0.051)	(0.088)	(0.101)
2010	-0.213	-0.246	-0.033	0.076	-0.012	-0.088	0.024	0.031	0.007
	(0.116)	(0.142)	(0.183)	(0.051)	(0.055)	(0.075)	(0.094)	(0.087)	(0.128)
2011	-0.278	-0.330	-0.052	0.122	0.135*	0.013	0.064	-0.008	-0.072
	(0.353)	(0.279)	(0.450)	(0.063)	(0.057)	(0.084)	(0.132)	(0.159)	(0.207)
2012	0.368	0.183	-0.185	0.082	-0.023	-0.105	-0.043	-0.046	-0.003
	(0.324)	(0.268)	(0.420)	(0.046)	(0.042)	(0.062)	(0.070)	(0.077)	(0.104)
2013	-0.145	-0.006	0.139	-0.039	0.036	0.075	0.030	-0.032	-0.063
	(0.153)	(0.160)	(0.221)	(0.032)	(0.037)	(0.049)	(0.052)	(0.071)	(0.088)
2014	-0.145	-0.247*	-0.102	0.000	0.041	0.041	-0.012	-0.081	-0.069
	(0.094)	(0.109)	(0.144)	(0.032)	(0.032)	(0.045)	(0.044)	(0.064)	(0.078)
2015	-0.412*	-0.036	0.376	0.020	0.123	0.103	0.004	-0.067	-0.071
	(0.171)	(0.207)	(0.268)	(0.052)	(0.065)	(0.083)	(0.070)	(0.093)	(0.116)
Simple ATTs	-0.146*	-0.115	-0.031	0.052**	0.042**	0.010	-0.060	-0.016	-0.043
	(0.066)	(0.067)	(0.095)	(0.017)	(0.016)	(0.023)	(0.032)	(0.024)	(0.040)

Notes. Standard errors in parentheses ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$). Estimates reflect estimates the group average treatment effect on the treated for each classification cohort as derived from a intent-to-treat student fixed effect model using the Callaway and Sant'Anna (2021) estimator. Estimates for students classified in an above-median suspension school and for those classified in a below-median suspension school, as well as the difference in effect sizes are displayed. Math and ELA scores are standardized to mean 0 and standard deviation of 1 and are available for students in grades 6-8. Attendance rates are the percent of days attended out of allowable/possible days and are available for all students. Students who are never classified with a disability are the omitted group.

Table 14: Group Average Treatment Effects of Classification: Attendance

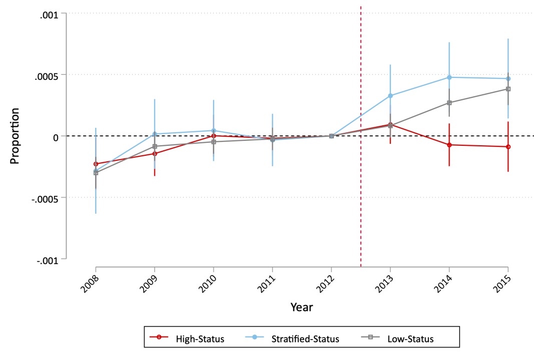
	Low-Status			Stratified Status			High Status		
Classification Cohort	*Below*	*Above*	*Difference*	*Below*	*Above*	*Difference*	*Below*	*Above*	*Difference*
2009	-8.950***	-10.293***	-1.343	-3.934***	-3.151***	0.783	-3.619**	-1.892	1.727
	(2.359)	(2.612)	(3.519)	(0.751)	(0.629)	(0.979)	(1.336)	(1.510)	(2.016)
2010	-12.472***	-10.552***	1.920	-5.754***	-3.333***	2.421	-3.472*	2.330	5.803*
	(2.855)	(2.643)	(3.891)	(0.921)	(0.866)	(1.264)	(1.716)	(1.752)	(2.453)
2011	-20.460**	-6.678	13.782	-4.997***	-3.371***	1.626	1.507	-2.520	-4.027
	(7.165)	(4.179)	(8.295)	(1.056)	(0.901)	(1.388)	(1.648)	(2.234)	(2.776)
2012	-12.808***	-7.303**	5.505	-2.755***	-2.592***	0.163	-1.191	1.120	2.311
	(3.432)	(2.808)	(4.434)	(0.741)	(0.677)	(1.004)	(1.074)	(1.159)	(1.580)
2013	-9.591**	-1.908	7.683	-3.197***	-2.292***	0.905	-1.356	0.267	1.623
	(3.456)	(2.157)	(4.074)	(0.492)	(0.466)	(0.678)	(0.814)	(1.213)	(1.460)
2014	-7.261**	-4.878*	2.383	-2.657***	-1.771***	0.887	-1.743*	-0.241	1.502
	(2.422)	(1.950)	(3.109)	(0.434)	(0.411)	(0.598)	(0.714)	(1.069)	(1.285)
2015	-6.924**	-2.431	4.493	-1.807*	-0.496	1.311	1.226	0.014	-1.212
	(2.151)	(2.326)	(3.168)	(0.710)	(0.628)	(0.948	(1.163)	(1.329)	(1.766)
Simple ATTs	-6.549***	-10.024***	3.475*	-2.537***	-3.615***	1.078**	0.024	-1.735***	1.759*
	(0.992)	(1.165)	(1.530)	(0.244)	(0.278)	(0.370)	(0.546)	(0.450)	(0.708)

Notes. Standard errors in parentheses ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$). Estimates reflect estimates the group average treatment effect on the treated for each classification cohort as derived from a intent-to-treat student fixed effect model using the Callaway and Sant'Anna (2021) estimator. Estimates for students classified in an above-median suspension school and for those classified in a below-median suspension school, as well as the difference in effect sizes are displayed. Math and ELA scores are standardized to mean 0 and standard deviation of 1 and are available for students in grades 6-8. Attendance rates are the percent of days attended out of allowable/possible days and are available for all students. Students who are never classified with a disability are the omitted group.

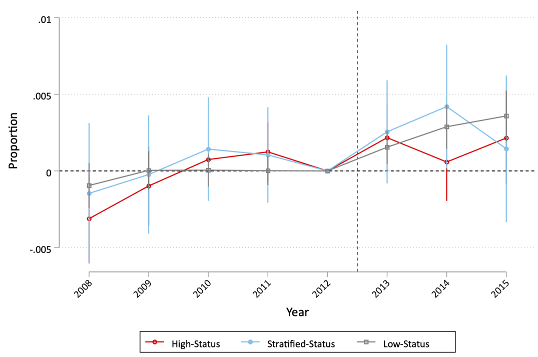
A Appendix

Figure A1: Alternative Definitions of Treatment

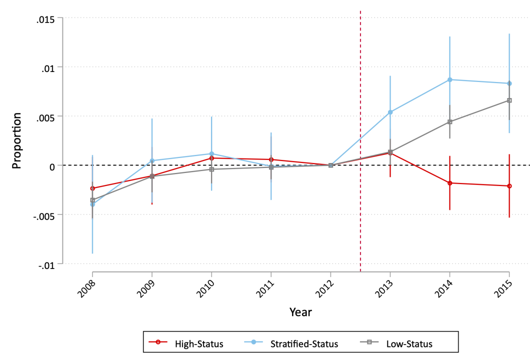
(a) Continuous Definition of Treatment



(b) 2nd Tercile Compared to 1st Tercile

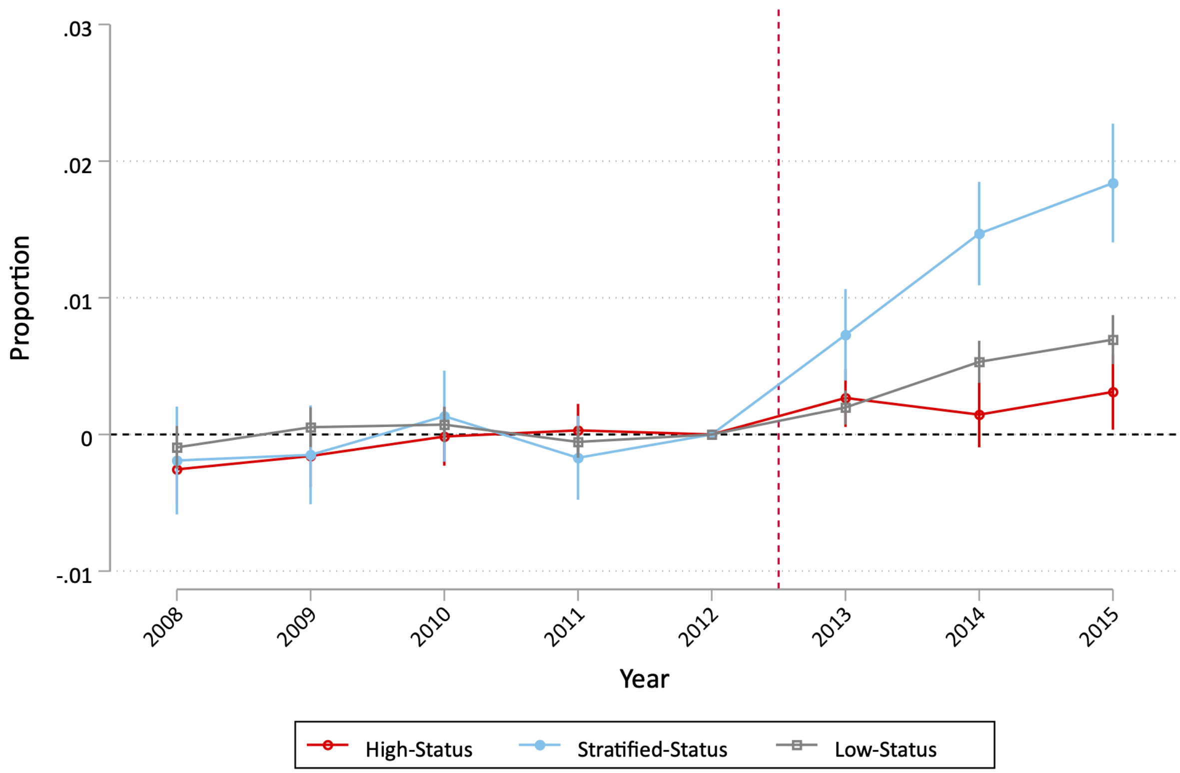


(c) 3rd Tercile Compared to 1st Tercile



Notes. Panel A compares differences in classification between the 2nd and 1st tercile of suspension rates relative to the 2012 difference; Panel B compares differences in classification between the 3rd and 1st tercile of suspension rates relative to the 2012 difference. Estimates in Panel C indicate the change in disability classification relative to 2012 for a 1 standard deviation increase in a school's 2012 suspension rate. Panels A and B use a school's tercile of 2011/12 suspension rates as the definition of treatment. All models use school-by-grade and year fixed effects and control for student gender, race, ELL status, FRPL eligibility, years present in NYCPS, 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.

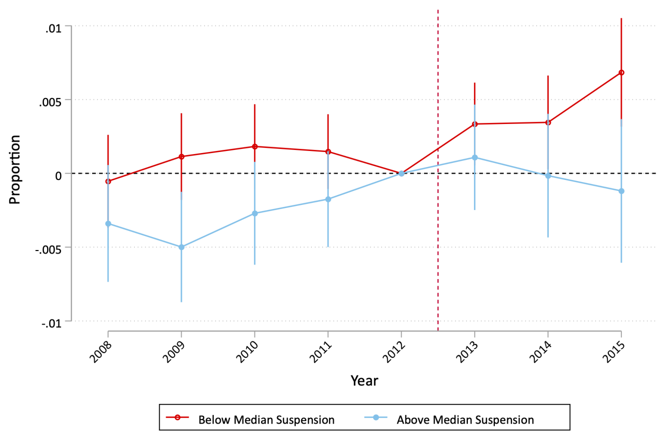
Figure A2: Heterogeneity by Race: Black−White/Asian



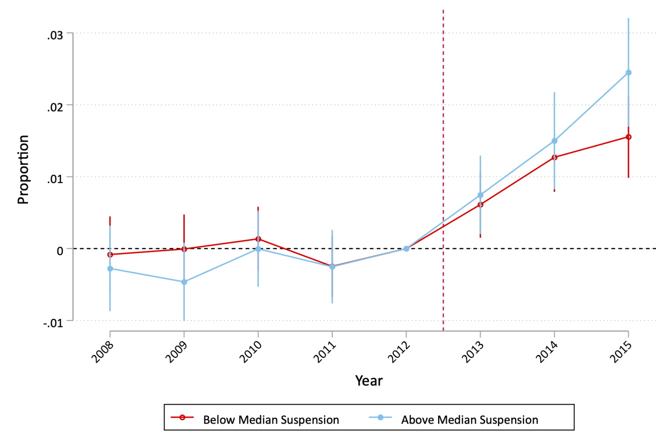
Notes. Estimates indicate the average change in classification rates for Black students relative to non-Black students relative to the difference in 2012. All models use school-by-grade and year fixed effects and control for student gender, race, ELL status, FRPL eligibility, years present in NYCPS, 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 provided.

Figure A3: Black−White/Asian Students in High and Low Suspension Schools

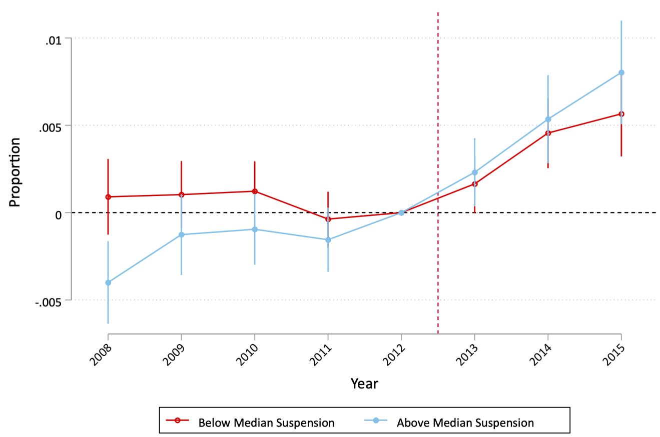
(a) High-Status



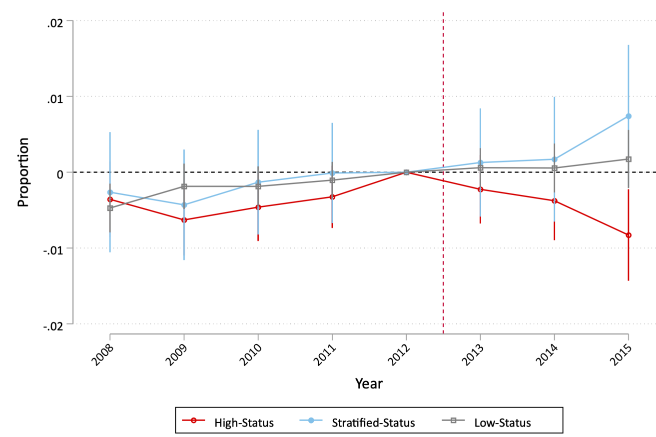
(b) Stratified Status



(c) Low-Status

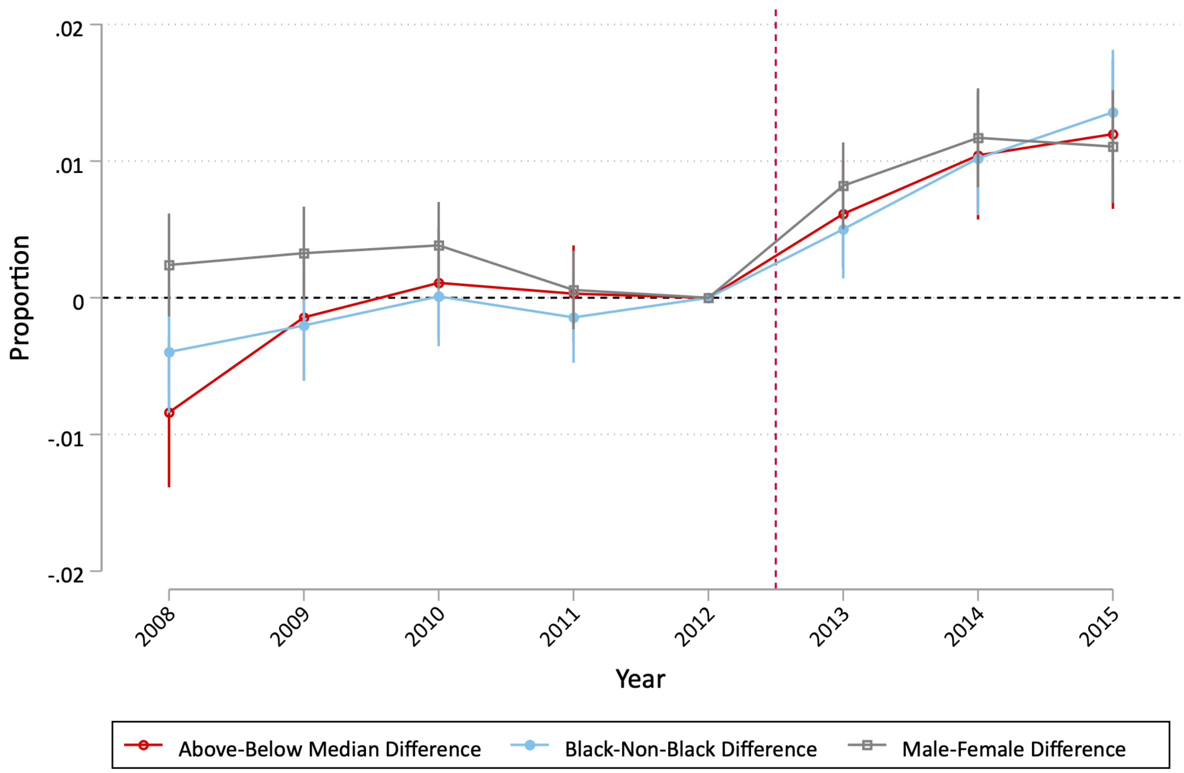


(d) Triple Difference



Notes. Estimates in Panels A-C presented represent the Black-White/Asian 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.

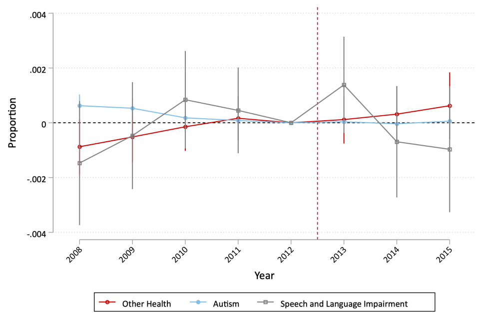
Figure A4: Non-Imputed SWD Indicator for Any Classification



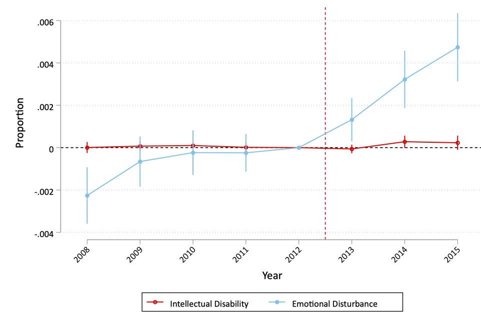
Notes. Estimates indicate the change in disability classification relative to 2012 for each treatment definition. All models use school-by-grade and year fixed effects and control for student gender, race, ELL status, FRPL eligibility, years present in NYCPS, mean pre-ban ELA and math test scores, a 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.

Figure A5: Estimates by Individual Disability Classification

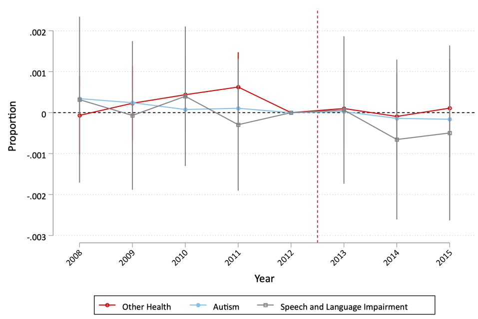
(a) Above−Below: High Status



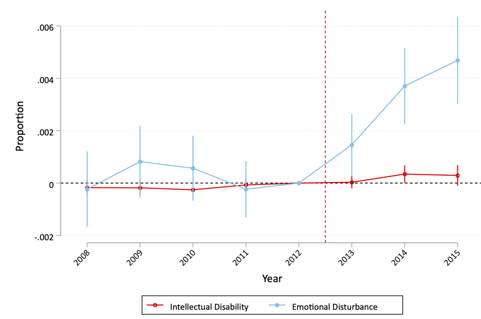
(b) Above−Below: Low Status



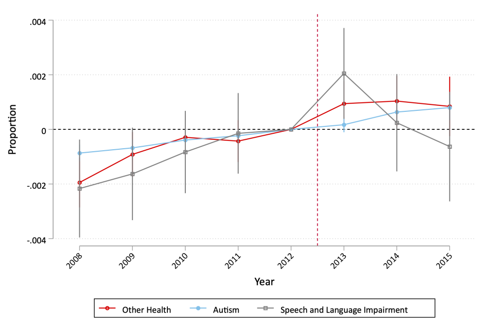
(c) Black−Non-Black: High Status



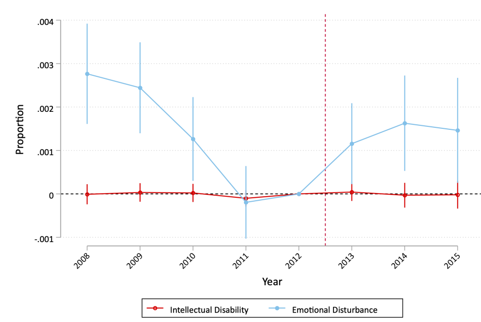
(d) Black−Non-Black: Low Status



(e) Male−Female: High Status

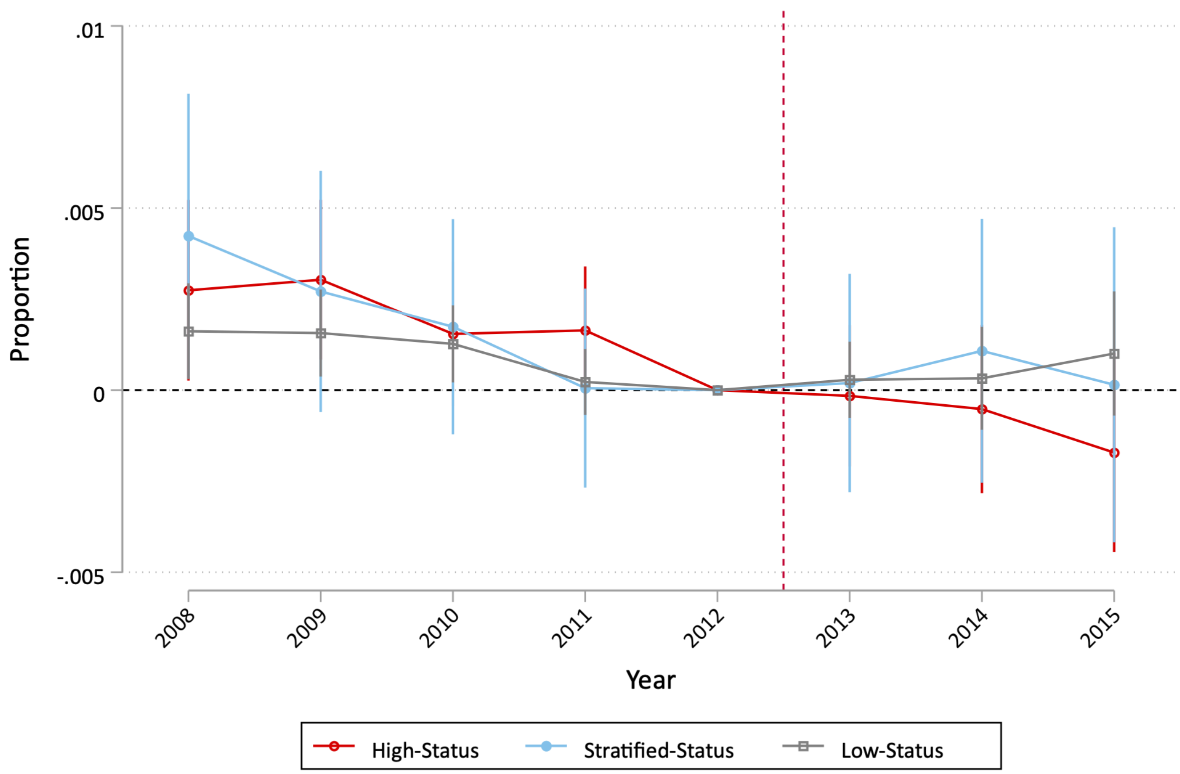


(f) Male−Female: Low Status



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 NYCPS, 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.

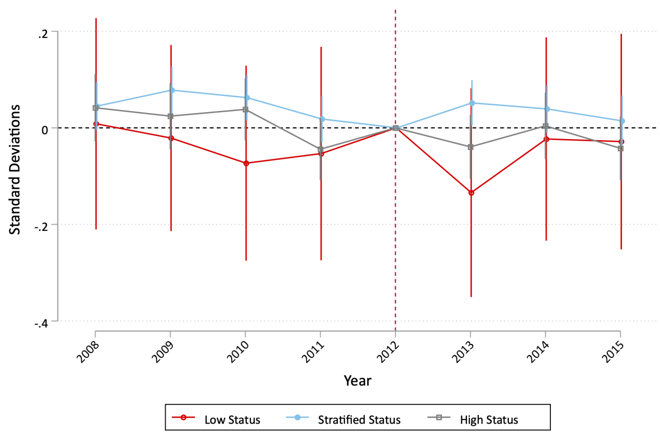
Figure A6: Differential Effects of aSPtS on Classification Behavior



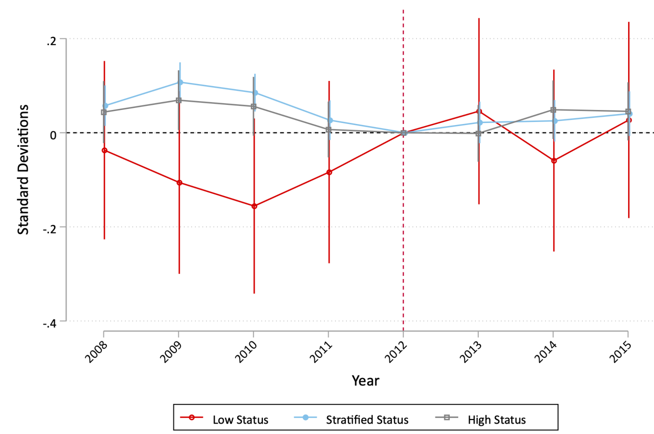
Notes. Estimates reflect the coefficients on interactions between year indicators and an indicator for being in a school with above-median rates of SWDs served in self-contained settings relative to 2012. 95% confidence intervals are displayed.

Figure A7: Differential Effects of the Suspension Ban on SWDs: Black Students

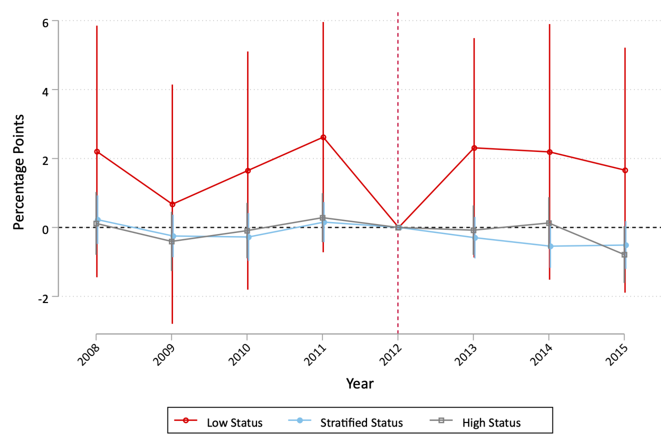
(a) Math Scores



(b) ELA Scores



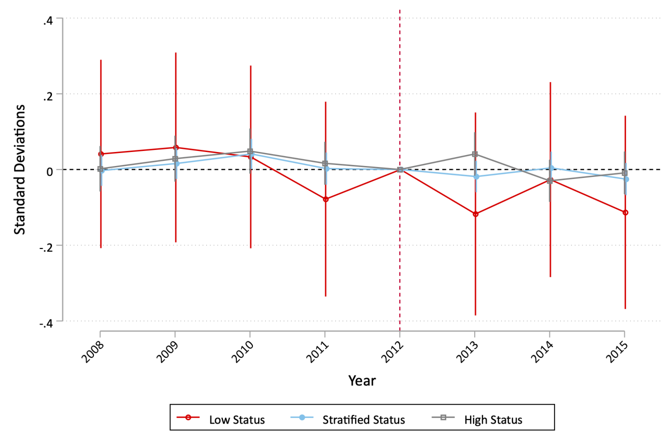
(c) Attendance



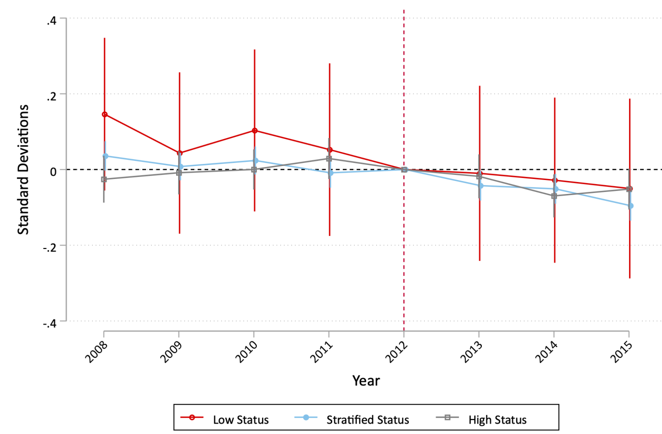
Notes. Estimates in Panels A-C presented represent the Black-non-Black gap in classification within a high or low-suspension school (above or below the 2011-12 district median suspension rate of 6 percent). Estimates in Panel D indicate the triple difference coefficient. 95% confidence intervals are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

Figure A8: Differential Effects of the Suspension Ban on SWDs: Male Students

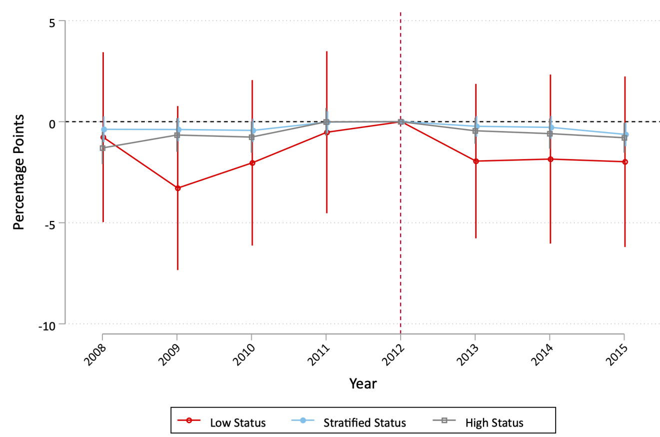
(a) Math Scores



(b) ELA Scores

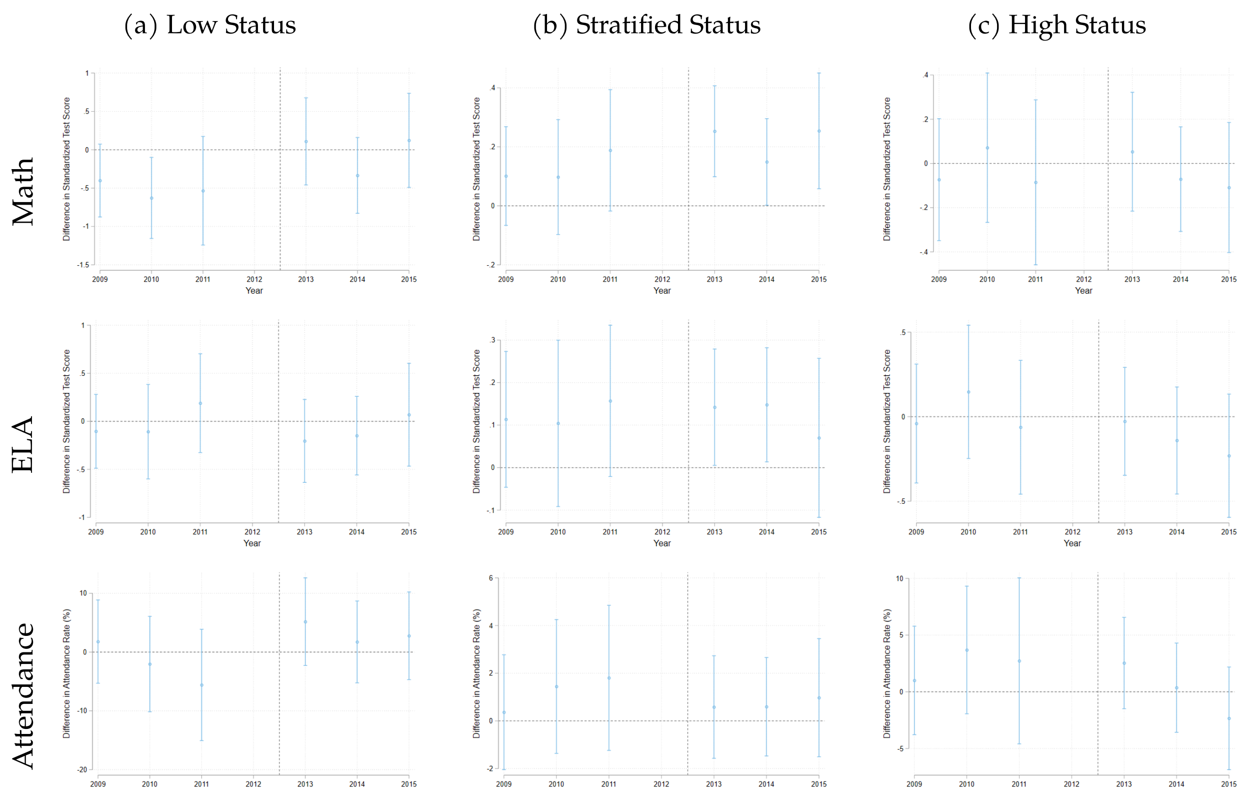


(c) Attendance



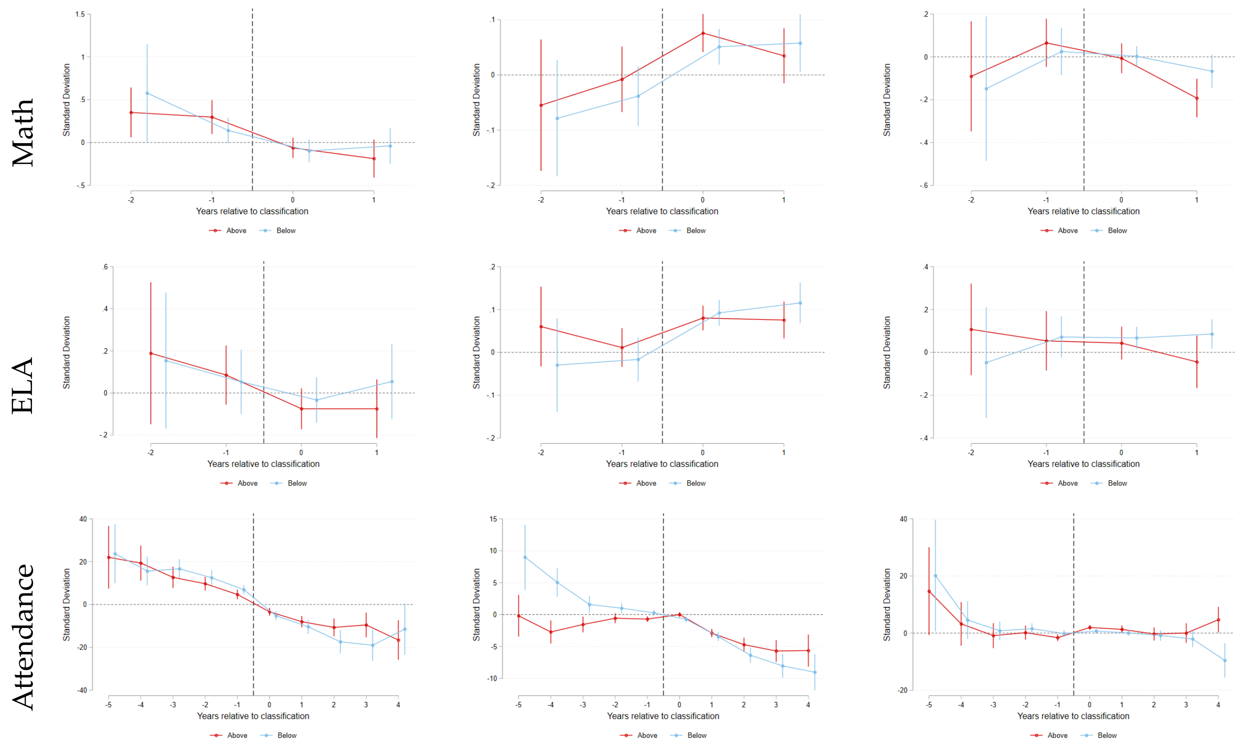
Notes. Estimates in Panels A-C presented represent the Black-non-Black gap in classification within a high or low-suspension school (above or below the 2011-12 district median suspension rate of 6 percent). Estimates in Panel D indicate the triple difference coefficient. 95% confidence intervals are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

Figure A9: Difference in Group Average Treatment Effects on the Treated between Above- and Below-Median Suspension Schools Relative to 2012: All Students



Notes. Estimates reflect the difference in the effect of receiving a subjective disability classification in a specific cohort as derived by an intent-to-treat student fixed effects model using the Callaway and Sant'Anna (2021) estimator relative to the effect size for those students classified in 2012. The comparison group in all estimates are students who never receive a disability classification. All models include controls for whether a student was held back and for a student's grade. The vertical dashed line indicates the year of the policy change (2012). 95% confidence intervals are displayed.

Figure A10: Event Study Estimates: Student Fixed Effect Models



(a) Low Status (b) Stratified Status (c) High Status

Notes. Estimates reflect event study estimates of classification across all cohorts as derived by an intent-to-treat student fixed effects model using the Callaway and Sant'Anna (2021) estimator. Separate estimates are provided for those classified in above-median suspension schools and for those classified in below-median suspension schools. The comparison group in all estimates are students who never receive a disability classification. All samples are restricted to SWDs that are only ever served in inclusive settings. The vertical dashed line indicates the time at which a student is classified. 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: Excluding Students Ever in District 75 Schools: Trends across Time and Difference-in-Difference Estimates

	Low-Status	**Stratified Status**	**High-Status**
Panel A. Trends across Time			
2013	0.001***	0.009***	0.007***
	(0.000)	(0.001)	(0.001)
2014	0.002***	0.016***	0.007***
	(0.000)	(0.001)	(0.001)
2015	0.002***	0.018***	0.010***
	(0.000)	(0.001)	(0.001)
Panel B. Above-Below Median Suspension			
2013	0.001*	0.008***	0.006***
	(0.000)	(0.001)	(0.001)
2014	0.002***	0.013***	0.008***
	(0.000)	(0.001)	(0.001)
2015	0.002***	0.016***	0.010***
	(0.000)	(0.001)	(0.001)
Above Median ×2013	0.001	0.003*	0.002
	(0.000)	(0.002)	(0.001)
Above Median ×2014	0.002***	0.007***	-0.000
	(0.001)	(0.002)	(0.001)
Above Median ×2015	0.003***	0.007***	0.000
	(0.001)	(0.002)	(0.001)
Panel C. Black-Non-Black			
2013	0.000	0.007***	0.007***
	(0.000)	(0.001)	(0.001)
2014	0.001***	0.013***	0.008***
	(0.000)	(0.001)	(0.001)
2015	0.001**	0.015***	0.010***
	(0.000)	(0.001)	(0.001)
Black ×2013	0.001	0.004*	-0.000
	(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	0.007***	0.005***
	(0.000)	(0.001)	(0.001)
2014	0.002***	0.011***	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	0.004**	0.003***
	(0.000)	(0.001)	(0.001)
Male ×2014	0.000	0.008***	0.002*
	(0.001)	(0.001)	(0.001)
Male ×2015	-0.000	0.008***	0.001
	(0.001)	(0.002)	(0.001)
Observations	2,556,971	2,556,971	2,556,971
Clusters	2,899	2,899	2,899

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 NYCPS, 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 who were ever enrolled in a District 75 school.

Table A4: Triple Difference Estimates: Black-White/Asian in Above Median Suspension Schools

	Low Status	**Stratified Status**	**High Status**
2013 $\times$ Above Median $\times$ Black	0.001	0.001	-0.002
	(0.001)	(0.004)	(0.002)
2014 $\times$ Above Median $\times$ Black	0.001	0.002	-0.004
	(0.002)	(0.004)	(0.003)
2015 $\times$ Above Median $\times$ Black	0.002	0.007	-0.008**
	(0.002)	(0.005)	(0.003)
Observations	1,542,126	1,542,126	1,542,126
Clusters	2,856	2,856	2,856

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 NYCPS, 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: Overlap between Schools with Above- and Below-Median Suspension Rates and SWDs in Self-Contained Settings

Panel A. Percent of Sample in Each Category

	Self-Contained	
	Below-Median	Above-Median
Suspension		
Below-Median	30.14	22.90
Above-Median	20.78	26.18

Panel B. Baseline Suspension Rates for Each Category

	Self-Contained	
Suspension		
Below-Median	3.08	3.61
Above-Median	13.10	12.16

Table A6: Changes in Characteristics Over Time between Above and Below-Median Suspension Schools

	Female	**FRPL**	**ELL**	**Black**	**Non-Structural Movers**
2013	-0.004*	-0.057***	-0.000	-0.004**	-0.001
	(0.002)	(0.006)	(0.001)	(0.001)	(0.001)
2014	-0.007***	-0.053***	-0.002	-0.007***	0.003**
	(0.002)	(0.006)	(0.001)	(0.002)	(0.001)
2015	-0.009***	-0.049***	-0.006***	-0.010***	0.000
	(0.002)	(0.007)	(0.001)	(0.002)	(0.001)
2013 $\times$ Above	0.001	-0.006	0.000	0.001	-0.001
	(0.003)	(0.009)	(0.002)	(0.002)	(0.001)
2014 $\times$ Above	-0.003	-0.004	0.000	-0.000	0.002
	(0.003)	(0.008)	(0.002)	(0.003)	(0.001)
2015 $\times$ Above	-0.007*	-0.009	0.002	-0.002	0.004*
	(0.003)	(0.009)	(0.002)	(0.003)	(0.002)
Observations	2,588,870	2,588,870	2,588,870	2,588,870	2,588,870

Notes. Standard errors in parentheses ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$) and are clustered at the school-grade level. No models include controls. 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 fixed effects are included in all models.

Table A7: Effects of the Suspension Ban on Disability Classification: Excluding Non-Structural Movers

	Low-Status	**Stratified Status**	**High-Status**
2013	0.001*	0.007***	0.006***
	(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)
2013 × Above	0.001*	0.003*	0.001
	(0.001)	(0.002)	(0.001)
2014 × Above	0.003***	0.007***	-0.001
	(0.001)	(0.002)	(0.001)
2015 × Above	0.005***	0.007**	-0.000
	(0.001)	(0.002)	(0.001)
Observations	2,508,626	2,508,626	2,508,626

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 NYCPS, 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%. All models exclude students who made a non-structural move (i.e., did not switch schools due to moving from middle to high school, that is, from grades 8 to 9).

Table A8: Differences in Attritors between Above and Below Median Suspension Schools

	Attritor	Female	FRPL	ELL	Black	Low Status	Stratified Status	High Status	Any Classification
2012	0.003**	-0.002	0.166***	0.001	-0.004**	0.000	0.000	0.001*	0.001
	(0.001)	(0.002)	(0.008)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)
2013	0.006***	-0.006**	0.108***	0.001	-0.009***	0.001**	0.007***	0.008***	0.017***
	(0.001)	(0.002)	(0.006)	(0.001)	(0.002)	(0.000)	(0.001)	(0.001)	(0.001)
2014	0.010***	-0.009***	0.112***	-0.001	-0.012***	0.002***	0.015***	0.011***	0.028***
	(0.002)	(0.002)	(0.006)	(0.001)	(0.002)	(0.000)	(0.001)	(0.001)	(0.002)
2012 × Above	0.001	-0.001	0.004	0.000	-0.001	-0.000	0.001	0.000	0.002
	(0.002)	(0.002)	(0.012)	(0.002)	(0.002)	(0.000)	(0.001)	(0.001)	(0.002)
2013 × Above	0.006*	-0.000	-0.003	-0.000	0.001	0.001	0.006**	0.002	0.009***
	(0.002)	(0.003)	(0.008)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
2014 × Above	0.012***	-0.005	-0.002	0.001	-0.001	0.003***	0.012***	0.002	0.017***
	(0.003)	(0.003)	(0.009)	(0.002)	(0.003)	(0.001)	(0.002)	(0.001)	(0.003)
2012 × Above × Attritor		-0.004	-0.004	0.007	0.006	0.003	0.002	-0.001	0.002
		(0.009)	(0.013)	(0.005)	(0.007)	(0.003)	(0.005)	(0.003)	(0.006)
2013 × Above × Attritor		0.002	0.003	0.010	0.006	0.002	0.001	0.002	0.003
		(0.009)	(0.011)	(0.005)	(0.008)	(0.003)	(0.007)	(0.004)	(0.009)
2014 × Above × Attritor		0.015	0.015	0.001	0.009	0.003	-0.001	-0.002	-0.003
		(0.009)	(0.011)	(0.005)	(0.008)	(0.003)	(0.007)	(0.004)	(0.009)
Observations	2,383,102	2,383,102	2,383,102	2,383,102	2,383,102	2,383,102	2,383,102	2,383,102	2,383,102

Notes. Standard errors in parentheses ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$) and are clustered at the school-grade level. No models include controls. Above-median indicates that a school's suspension rate in the 2011-12 AY was greater than the district median of 6%. Attritor is an indicator for a student that leaves the sample the following year for an unobservable reason. As such, all estimates are anchored to 2010/2011—students who leave in the year prior to the policy change (2011/12)—and extend to 2013/14—students who leave in 2014/15. Estimation sample therefore excludes the 2014/15 academic year, as reflected by the smaller sample size relative to other estimations. School by grade fixed effects are included in all models.

Table A9: 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.

Table A10: Changes in Outcomes for SWDs: Male Students Relative to Female Students

	Any	Low	Stratified	High
Panel A. Math				
Disability × Male × 2013	0.008	-0.117	-0.018	0.041
	(0.018)	(0.137)	(0.021)	(0.029)
Disability × Male × 2014	-0.003	-0.027	0.004	-0.030
	(0.017)	(0.131)	(0.022)	(0.028)
Disability × Male × 2015	-0.021	-0.113	-0.025	-0.009
	(0.017)	(0.130)	(0.022)	(0.029)
Observations	1,235,536	1,106,716	1,185,703	1,149,029
Panel B. ELA				
Disability × Male × 2013	-0.023	-0.010	-0.043*	-0.018
	(0.016)	(0.118)	(0.020)	(0.030)
Disability × Male × 2014	-0.050**	-0.028	-0.051*	-0.070*
	(0.016)	(0.111)	(0.020)	(0.029)
Disability × Male × 2015	-0.078***	-0.050	-0.095***	-0.052
	(0.016)	(0.121)	(0.020)	(0.028)
Observations	1,230,062	1,100,705	1,179,946	1,143,233
Panel C. Attendance Rates				
Disability × Male × 2013	-0.117	-1.950	-0.229	-0.448
	(0.208)	(1.948)	(0.261)	(0.330)
Disability × Male × 2014	-0.241	-1.849	-0.280	-0.590
	(0.237)	(2.132)	(0.290)	(0.380)
Disability × Male × 2015	-0.439	-1.980	-0.631*	-0.794*
	(0.241)	(2.151)	(0.296)	(0.378)
Observations	3,415,156	3,128,624	3,320,625	3,201,673

Notes. Standard errors in parentheses ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$) and are clustered at the school-grade level. 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.

Table A11: Changes in Outcomes for SWDs: Black Students Relative to Non-Black Students

	Any	Low	Stratified	High
Panel A. Math				
Disability × Black × 2013	0.011	-0.134	0.052*	-0.039
	(0.020)	(0.110)	(0.024)	(0.034)
Disability × Black × 2014	0.020	-0.023	0.039	0.004
	(0.021)	(0.107)	(0.025)	(0.035)
Disability × Black × 2015	-0.012	-0.028	0.015	-0.042
	(0.022)	(0.114)	(0.027)	(0.033)
Observations	1,235,536	1,106,716	1,185,703	1,149,029
Panel B. ELA				
Disability × Black × 2013	0.011	0.046	0.022	-0.001
	(0.018)	(0.101)	(0.023)	(0.031)
Disability × Black × 2014	0.020	-0.059	0.025	0.049
	(0.018)	(0.099)	(0.023)	(0.032)
Disability × Black × 2015	0.033	0.027	0.041	0.045
	(0.019)	(0.106)	(0.024)	(0.031)
Observations	1,230,062	1,100,705	1,179,946	1,143,233
Panel C. Attendance Rates				
Disability × Black × 2013	-0.171	2.310	-0.296	-0.078
	(0.244)	(1.623)	(0.306)	(0.366)
Disability × Black × 2014	-0.485	2.192	-0.542	0.130
	(0.261)	(1.891)	(0.321)	(0.382)
Disability × Black × 2015	-0.667*	1.662	-0.510	-0.785
	(0.284)	(1.812)	(0.353)	(0.420)
Observations	3,415,156	3,128,624	3,320,625	3,201,673

Notes. Standard errors in parentheses ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$) and are clustered at the school-grade level. 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.

Table A12: Differential Effects by School Suspension Rate and SWDs in Self-Contained Settings

	Any	Low	Stratified	High
Panel A. Math Scores				
Dis. × Suspension × Self-Contained × 2013	0.000	-0.013	0.028	-0.033
	(0.041)	(0.213)	(0.049)	(0.064)
Dis. × Suspension × Self-Contained × 2014	-0.041	-0.244	-0.012	-0.045
	(0.046)	(0.207)	(0.054)	(0.065)
Dis. × Suspension × Self-Contained × 2015	-0.010	-0.121	0.050	-0.006
	(0.044)	(0.210)	(0.055)	(0.063)
Observations	1,226,528	1,098,573	1,177,045	1,140,588
Panel B. ELA Scores				
Dis. × Suspension × Self-Contained × 2013	0.033	0.102	0.039	0.015
	(0.037)	(0.197)	(0.044)	(0.060)
Dis. × Suspension × Self-Contained × 2014	0.042	0.032	0.053	-0.006
	(0.038)	(0.198)	(0.043)	(0.060)
Dis. × Suspension × Self-Contained × 2015	0.051	0.033	0.068	0.020
	(0.040)	(0.213)	(0.050)	(0.058)
Observations	1,222,177	1,093,683	1,172,410	1,135,914
Panel C. Attendance Rates				
Dis. × Suspension × Self-Contained × 2013	-0.529	-0.477	-0.003	-0.922
	(0.479)	(3.307)	(0.603)	(0.639)
Dis. × Suspension × Self-Contained × 2013	-0.529	-0.273	-0.208	-1.069
	(0.479)	(3.333)	(0.598)	(0.641)
Dis. × Suspension × Self-Contained × 2014	0.483	-2.534	1.128	-0.618
	(0.535)	(3.673)	(0.676)	(0.707)
Dis. × Suspension × Self-Contained × 2015	0.228	3.736	0.497	-1.148
	(0.572)	(3.835)	(0.738)	(0.755)
Observations	3,388,033	3,101,551	3,293,522	3,174,594

Notes. Standard errors in parentheses ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$) and are clustered at the school-grade level. "Suspension" is an indicator for being in a school with a pre-ban above-median suspension rate; "Dis." is an indicator for either any, low, stratified, or high-status disability and equal to 0 if a student has no disability; "Self-Contained" is an indicator for being in a school with pre-ban above-median rate of students served in self-contained settings in 2012.

Table A13: Difference between Pre-Classification Standardized Test Scores for Students Classified with a Low-Status Disability Pre- and Post-Ban

	Pre-Ban	**Post-Ban**	**Difference**
ELA			
Below-Median	-0.923	-0.899	-0.024
Above-Median	-0.981	-1.094	0.113***
Math			
Below-Median	-1.087	-0.910	-0.177***
Above-Median	-1.139	-1.103	-0.036

Notes. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Estimates indicate the average pre-classification test scores for students classified with a low-status disability prior to the ban and after the ban. Below-median indicates that a student was in a school with below-median suspension rates; above-median indicates that a student was in a school with above-median suspension rates