When Does Descriptive Representation Reduce Racial Disparities?

Evidence from California School Board Elections

Igor Geyn*

April 11, 2026

Abstract

Does electing minority representatives reduce racial disparities in public services? I exploit staggered transitions from at-large to single-member district (SMD) school board elections induced by California Voting Rights Act litigation to estimate causal effects on racial discipline gaps. Contrary to expectations, SMD adoption increases Black-White suspension gaps by 2.7 percentage points (Sun-Abraham; $p = .003$) while leaving Hispanic-White gaps and total suspension rates unchanged. I investigate this puzzle by showing that SMD elections primarily increase Hispanic, not Black, board representation—and that discipline disparities decline only when the race of elected representatives matches the affected student group. Where Black board members are elected, Black-White gaps are attenuated (-34 points per unit board share, approximately -1 to -2 pp per standard deviation of Black board share; $p = .005$); where Black teachers are present, gaps are similarly moderated (-82 to -96 points per unit

*Independent researcher. Ph.D., Political Science, UCLA. Email: igorgeyn@gmail.com. For helpful comments and suggestions, the author thanks Julia Payson, Dan Thompson, Chris Tausanovitch, Joshua Ferrer, Chris Palmisano, Morgan Spangler, Graham Straus, and seminar participants at the UCLA American Politics weekly research group. All errors are my own.

teacher share; $p < .001$). However, the board composition moderation is sensitive to the BISG race-prediction specification and does not survive Benjamini-Yekutieli multiple-testing correction, whereas the teacher diversity channel does. These results are consistent with a race-match condition for descriptive representation: "minority" representation is not fungible across racial groups. Because the moderation evidence rests on post-treatment composition, which is endogenous, I interpret the race-match pattern as suggestive rather than causally identified.

1 Introduction

Over the past two decades, more than two hundred California school districts have been compelled to abandon at-large elections in favor of single-member district (SMD) systems for their governing boards. This wave of electoral reform was driven by the California Voting Rights Act (CVRA) of 2001, which enables legal challenges to at-large elections that dilute minority voting power. Because at-large elections require candidates to win jurisdiction-wide majorities, they can systematically disadvantage geographically concentrated minority communities. SMD elections, by contrast, create smaller electoral districts in which minority voters may constitute majorities, thereby increasing the likelihood that minority-preferred candidates win seats. The CVRA reform wave thus provides a rare opportunity to study whether changes in the racial composition of elected officials translate into changes in the policies and outcomes those officials oversee.

The theoretical prediction is straightforward. A large literature on descriptive representation argues that elected officials who share the racial or ethnic background of their constituents are more likely to advance those constituents' interests (Pitkin 1967; Mansbridge 1999; Meier and England 1984). In education specifically, minority school board members have been linked to more equitable hiring practices, more favorable resource allocation to minority-serving schools, and reduced racial disparities in discipline (Meier, Stewart, and England 1989; Leal, Martinez-Ebers, and Meier 2004; Meier and Rutherford 2016). Racial disparities in school discipline—particularly the disproportionate suspension of Black students—are among the most persistent and consequential inequities in American education (Skiba et al. 2002, 2011; Morris and Perry 2016). If SMD elections increase minority representation, and minority representatives reduce discipline disparities, then electoral reform should narrow racial gaps in suspensions.

This paper tests that prediction and finds that it fails, but in an informative way.

Using a staggered difference-in-differences design with 230 treated and 558 control school districts observed from 2011 to 2024, I estimate that SMD adoption increases the

Black-White suspension rate gap by approximately 2.7 percentage points (Sun-Abraham estimator; SE = 0.92; $p = .003$). Callaway-Sant'Anna confirms the direction and significance (+2.17; SE = 0.87; $p = .013$), and even the potentially biased two-way fixed effects estimator yields +1.28 (SE = 0.67; $p = .058$). Borusyak-Jaravel-Spiess imputation yields a smaller, insignificant estimate of +1.20 (SE = 0.74; $p = .102$), likely reflecting sensitivity to the imputation model's extrapolation across the many treatment cohorts in this setting. The Hispanic-White suspension gap is essentially unaffected, and total suspension rates show no significant change. The effect is concentrated in Black suspension rates, which increase by 1.8 percentage points after SMD adoption (TWFE; race-specific SA estimates are qualitatively similar), while White and Hispanic rates remain flat.

These results present a puzzle. Why would an electoral reform designed to empower minority communities worsen the racial discipline gap for Black students? The puzzle is sharpened by recent evidence that school board incumbents face no electoral accountability for minority student outcomes—only White student performance predicts reelection (Flavin and Hartney 2017). If voters do not punish boards for racial disparities, then changing who sits on the board is the primary lever for change, and the specific racial composition of that change becomes the binding constraint.

I investigate this puzzle by documenting what I call the *race-match condition* for descriptive representation. The key insight is that the CVRA reform wave in California primarily increased Hispanic, not Black, board representation. Among elected school board members in treated districts, the Hispanic share rose from 26.7% before SMD adoption to 36.5% afterward—an increase of nearly 10 percentage points. The Black share, by contrast, actually declined slightly, from 6.6% to 5.9%. This pattern reflects California's demography: Latino communities are the primary beneficiaries of CVRA enforcement because they are the largest minority group with sufficient residential concentration to form SMD majorities in most districts. Only 17 of the 230 treated districts in my sample gained a predicted-Black board member after switching to SMD elections; 163 gained a predicted-Hispanic member.

The race-match condition is confirmed through moderation analysis. When I interact the SMD treatment indicator with the cumulative Black share of the school board, I find a large and significant negative coefficient: -36.3 points per unit share (SE $= 11.9$; $p = .002$). In substantive terms, a one-standard-deviation increase in Black board representation is associated with a 3–4 percentage point reduction in the Black-White suspension gap. The analogous mismatch test—interacting the treatment indicator with Hispanic board share, with the BW gap as outcome—yields a null result (-2.4; $p = .48$). The pattern is similarly race-specific for Hispanic-White gaps: neither Hispanic board representation (-1.6; $p = .23$) nor Black board representation ($+0.4$; $p = .93$) significantly moderates the Hispanic-White gap.

A parallel mechanism operates through the teacher workforce. SMD adoption increases minority teacher shares by 1.5 percentage points (Sun-Abraham; $p = .029$), with the gain concentrated among Hispanic teachers ($+1.4$ pp; $p = .001$). Black teacher representation increases modestly under TWFE ($+0.3$ pp; $p = .003$). Black-White suspension gaps decline substantially in the presence of Black teachers (-82.0 pp per unit share, i.e., a 1 percentage-point increase in Black teacher share is associated with a 0.82 pp reduction in the BW gap; $p = .0003$; lagged: -95.8; $p = .0001$). Hispanic teachers have no significant effect on Black-White gaps ($+1.3$; $p = .70$), confirming the race-specificity of the mechanism.

These findings make three contributions. First, I provide the first causal estimates linking electoral reform to school discipline outcomes. Prior work has established that SMD elections increase minority representation (Abott and Magazinnik 2020; Trebbi, Aghion, and Alesina 2008; Trounstine and Valdini 2008), that minority representation is associated with more equitable bureaucratic outcomes (Meier, Stewart, and England 1989; Kogan, Lavertu, and Peskowitz 2021), and that same-race board members causally improve spending and achievement in their group's schools (Fischer 2023). No study has connected electoral reform to discipline—the domain where racial disparities are among the largest and most consequential for affected students' life trajectories. Second, I document a race-match pattern that appears to act as a boundary condition for theories of descriptive representation. The claim that "minority" representation reduces racial disparities appears to hold only when the representative's race matches that of the affected group, challenging the common practice of aggregating racial minorities into a single category for representation analysis. I interpret this pattern as suggestive rather than causally identified, since post-treatment board composition is endogenous to district characteristics that may independently affect discipline. Third, I show that the downstream effects of voting rights enforcement depend on local racial demography. Where the primary beneficiary group (here, Hispanic voters) differs from the group experiencing the targeted disparity (here, Black students), electoral reform alone may not reduce—and could even inadvertently increase—racial inequities.

The remainder of the paper proceeds as follows. Section 2 reviews the CVRA, the literature on electoral institutions and representation, and the theory of race-matched descriptive representation. Section 3 describes the data sources, sample construction, and identification strategy. Section 4 presents the main results, event studies, and robustness checks. Section 5 documents the race-match channel through board composition and teacher diversity. Section 6 discusses implications and limitations.

2 Background and Theory

2.1 The California Voting Rights Act

The California Voting Rights Act of 2001 (CVRA; California Elections Code §14025–14032) prohibits the use of at-large election methods when such methods impair the ability of members of a protected class to elect candidates of their choice or to influence the outcome of an election. Unlike Section 2 of the federal Voting Rights Act, the CVRA does not require plaintiffs to demonstrate that a compact, geographically defined minority community could constitute a majority in a hypothetical single-member district (Collingwood and Long 2021).

This lower evidentiary standard makes CVRA challenges substantially easier to mount than federal challenges.

A "safe harbor" provision added to the Elections Code in 2014 (Section 10010) further accelerated transitions. Under the safe harbor, a jurisdiction that passes a resolution of intent to transition to district-based elections triggers a 90-day protected window (extendable by mutual agreement) during which the jurisdiction can adopt a redistricting plan. If the jurisdiction completes the transition within this window, its reimbursement obligation to the prospective plaintiff is capped at approximately $30,000 (adjusted annually for inflation). By contrast, jurisdictions that resist and lose a contested CVRA lawsuit face attorneys' fees that routinely exceed $1 million. This asymmetry created strong incentives for districts to settle voluntarily upon receiving demand letters rather than risk litigation.

The litigation process follows a characteristic pattern. A small number of law firms specializing in CVRA enforcement identify target jurisdictions based on demographic analysis—specifically, whether a jurisdiction has sufficient residential segregation and Latino citizen voting-age population to support a viable SMD plan. The firm then sends a demand letter to the jurisdiction's governing body, notifying it of the legal vulnerability and offering to negotiate a transition to district-based elections. Faced with the near-certainty of losing a CVRA challenge and the prospect of paying the plaintiff's legal fees, the vast majority of jurisdictions settle by adopting SMD elections (Abott and Magazinnik 2020; Collingwood and Long 2021).

For causal identification, the critical feature of CVRA enforcement is that the *timing* of legal challenges is driven primarily by the capacity constraints and strategic choices of a small number of plaintiff attorneys, not by changes in the targeted jurisdictions' policies or demographic trajectories. While the *selection* of targets is related to district demographics—larger, more diverse districts with greater residential segregation are more likely to be targeted—the *sequencing* of challenges within this eligible pool is plausibly exogenous to school discipline outcomes. Some demographically similar districts were challenged in 2012 while others were not challenged until 2018 or later, a pattern driven by lawyers' caseload capacity rather than changes in the districts themselves.

2.2 Electoral Structure and Descriptive Representation

A foundational literature documents that at-large elections systematically reduce minority representation relative to single-member district systems (Engstrom and McDonald 1981; Davidson and Grofman 1994; Trebbi, Aghion, and Alesina 2008). The mechanism is well understood: at-large elections require jurisdiction-wide majorities, so under racially polarized voting, minority-preferred candidates are consistently outvoted. SMD elections create smaller electoral units in which minority communities can constitute majorities, enabling the election of minority-preferred candidates.

Abott and Magazinnik (2020) provide the most directly relevant evidence, studying the same CVRA-induced transitions I examine here. Using data on California school boards and city councils from 2002 to 2016, they find that switching from at-large to SMD elections increases Latino representation by approximately 4.5 percentage points (from a base of roughly 16%), but has a smaller and statistically insignificant effect on Black and Asian representation. Trounstine and Valdini (2008) similarly show that SMD systems are associated with greater diversity on city councils, though the effects vary by group and context. Marschall, Ruhil, and Shah (2010) demonstrate that electoral structure interacts with minority population size: SMD elections increase Black representation primarily where Black populations are sufficiently large and geographically concentrated.

These findings are central to the present paper's argument. If SMD elections disproportionately increase Hispanic representation while leaving Black representation largely unchanged—as Abott and Magazinnik (2020) document—then the downstream effects on Black-specific outcomes may differ from the effects on Hispanic-specific outcomes.

2.3 Descriptive Representation and Policy Outcomes

The theoretical case for descriptive representation is that elected officials who share constituents' demographic characteristics are more likely to understand and prioritize their concerns (Pitkin 1967; Mansbridge 1999). In education governance, Meier and England (1984); Meier, Stewart, and England (1989) show that Black school board representation is associated with more Black teachers, fewer Black students in special education, and lower Black suspension rates—with effects mediated by "bureaucratic representation" (hiring minority administrators and teachers who exercise discretion more equitably). Kogan, Lavertu, and Peskowitz (2021) provide modern causal evidence that minority board members reduce the Black-White achievement gap (≈ 0.1 SD) and reduce disproportionate discipline policies.

A critical question is why at-large boards tolerate racial disparities in the first place. Flavin and Hartney (2017) find that school board incumbents in California are electorally rewarded for White student achievement but face no accountability for Black student outcomes (formal equality test $p = .03$). Even in majority-minority districts, only White outcomes drive incumbent reelection. This structural accountability vacuum means that absent descriptive representation, there is no electoral mechanism to translate racial equity concerns into board action. Payson (2017) documents a compounding factor: performance accountability is limited to higher-turnout presidential-year elections, while the off-cycle elections that characterize most school board races are dominated by organized interests.

Fischer (2023) provides the most direct causal evidence for race-specific representation effects. Using a ballot-order instrument in California school board elections, he shows that electing a Hispanic board member causally increases per-pupil spending at majority-Hispanic schools (+\$71; +0.04 SD math achievement) while leaving other schools unchanged. The effect is race-specific: Hispanic board members benefit Hispanic-serving schools, not schools serving other minority groups—directly prefiguring the race-match pattern I document in discipline.

The teacher race-match literature provides complementary evidence. Dee (2005) finds that own-race teachers reduce the probability of being disciplined; Gershenson et al. (2022) show that exposure to a same-race teacher in elementary school reduces high school dropout by 39% for Black males; and Lindsay and Hart (2017) finds that Black students exposed to same-race teachers are significantly less likely to receive exclusionary discipline.

2.4 The Race-Match Condition

The existing literature on descriptive representation typically treats "minority" representation as a single construct. Studies examine whether minority board members reduce discipline disparities for minority students, or whether diverse teacher workforces improve outcomes for students of color broadly. This framing implicitly assumes a degree of cross-racial solidarity: that a Hispanic board member will advocate for Black students, or that an increase in the overall minority share of the teaching force will benefit Black students specifically.

I propose an alternative framing: the *race-match condition*. This condition states that the beneficial effects of descriptive representation are specific to the racial group represented. A Hispanic board member may prioritize the concerns of Hispanic families—including bilingual education, immigration-related anxieties, and culturally responsive pedagogy—without necessarily attending to the distinct concerns of Black families, such as racial profiling in discipline, culturally biased behavioral expectations, and the school-to-prison pipeline. The race-match condition does not require racial animosity or indifference; it simply reflects the fact that different racial groups face different challenges, and elected officials are most effective advocates for the specific communities from which they are drawn.

Existing evidence supports this framing. Fischer's (2023) finding that Hispanic board members increase spending at Hispanic-serving schools but not at other schools is a direct empirical precedent for race-specific representation effects in the spending domain. The accountability evidence in Flavin and Hartney (2017) provides a mechanism: under SMD elections, Hispanic board members face electoral accountability from Hispanic voters who may now constitute the median voter in their trustee area, but they have no more electoral incentive to attend to Black student outcomes than the White incumbents they replaced. The accountability deficit that Flavin and Hartney document under at-large elections is thus not eliminated by SMD reform—it is relocated, persisting for whichever racial groups remain too small or too dispersed to anchor a trustee-area majority.

The race-match condition generates a prediction for the present setting. If SMD elections primarily increase Hispanic representation (as Abott and Magazinnik 2020 show), then effects on discipline could appear in Hispanic-specific outcomes—though this prediction is conditional on there being a pre-existing Hispanic-White gap large enough to reduce. In the data, the baseline HW gap is close to zero (-0.10 pp in the full sample), which may limit the scope for improvement regardless of representation gains. The Black-White gap should be unaffected or could even worsen if, for example, increased attention to Hispanic students' needs diverts administrative resources from Black students' concerns. The average effect across all districts will reflect the composition of treatment effects: a weighted average of the (negative, gap-reducing) effects in districts that gained Black representation and the (null or positive, gap-widening) effects in districts that gained only Hispanic representation. If the latter group vastly outnumbers the former—as California's demography ensures—then the average effect will be dominated by null or adverse effects on the Black-White gap.

3 Data and Research Design

3.1 Data Sources

I combine data from four primary sources. *Election data* come from the California Elections Data Archive (CEDA), which provides candidate-level information on school board elections from 1996 through 2023, including candidate names, vote totals, incumbency status, and whether each election was conducted at-large or by trustee area (California Elections Data Archive 2024). The trustee area field is the basis for identifying SMD elections.

Discipline data come from the California Department of Education (CDE), which publishes annual district-level counts of suspensions and expulsions disaggregated by race/ethnicity for the 2010–11 through 2023–24 academic years (California Department of Education 2024). I calculate race-specific suspension rates (suspensions per 100 enrolled students) for Black, Hispanic, and White students, and construct two primary outcome variables: the Black-White suspension rate gap and the Hispanic-White suspension rate gap, each measured in percentage points.

Teacher workforce data also come from the CDE, which reports district-level staff demographics disaggregated by race/ethnicity for the same period. I calculate the share of teachers in each district who are Black, Hispanic, and minority (non-White) in each year.

Validation data come from the Lawyers' Committee for Civil Rights of the San Francisco Bay Area (LCCR), which maintains a database of 102 California school districts that transitioned to trustee-area elections through November 2013 (Lawyers' Committee for Civil Rights of the San Francisco Bay Area 2014). I use this database to validate and correct the CEDA-based treatment identification.

To predict the race/ethnicity of school board candidates—who are not asked to self-report race—I apply Bayesian Improved Surname Geocoding (BISG) using the `wru` package (Imai and Khanna 2016). BISG combines surname-based race probabilities with county-level demographic data from the Census to produce posterior probabilities of each candidate belonging to each racial/ethnic group. Approximately 79% of candidates have a surname that matches the Census surname list and receive full BISG predictions; the remaining 21% do not match the list and receive fallback predictions based on the county racial distribution alone.

3.2 Treatment Identification

Treatment—the transition from at-large to SMD elections—is identified from the CEDA `area` field. A district-year observation is coded as "SMD" if any election in that district in that year includes a non-missing trustee area identifier. The first year in which a district holds an SMD election is coded as the treatment year. Districts that always used at-large elections during the sample period serve as controls; districts that always used SMD elections are excluded.

I validate this identification against the LCCR database, which provides an independent record of CVRA-induced transitions. Of the 102 districts in the LCCR database, I match 93 to CEDA records using fuzzy string matching within county. This validation revealed 13 districts that were listed in the LCCR database as having transitioned but were not captured by the CEDA-based identification, typically because their first SMD election fell outside the CEDA sample window or because the CEDA area field was inconsistently recorded. I incorporate these 13 corrections, bringing the total treated districts from 244 to 257 in the full CEDA sample (230 after restricting to districts that merge with CDE discipline data).

The final treatment sample consists of 230 treated districts distributed across twenty cohorts spanning 1999–2023 (Figure 1). The bulk of the identifying variation comes from eight core CVRA-era cohorts (2012–2020, 156 districts), with the largest waves in 2018 (34) and 2020 (44). The remaining 74 treated districts fall in edge cohorts with limited estimation windows.[1] The 558 control districts maintained at-large elections throughout the sample period.

3.3 Outcome Variables

The primary outcomes are the Black-White suspension rate gap and the Hispanic-White suspension rate gap, measured as the difference in suspensions per 100 enrolled students between each minority group and White students. Racial gaps are the natural estimand because the theoretical prediction concerns disparities, not discipline levels per se: descriptive

[1]Twenty-two districts adopted SMD elections before the 2011–2024 analysis window (cohorts 1999–2010) and are observed only in post-treatment years. Under the Sun-Abraham and Callaway-Sant'Anna estimators, these districts are absorbed by district fixed effects and do not contribute to the ATT. The 2022 cohort (51 districts) has at most two post-treatment years; leave-one-cohort-out analysis (Section 4.4) confirms that excluding it strengthens the BW gap estimate slightly (+3.17 vs. baseline +2.74). One district adopted in 2023.

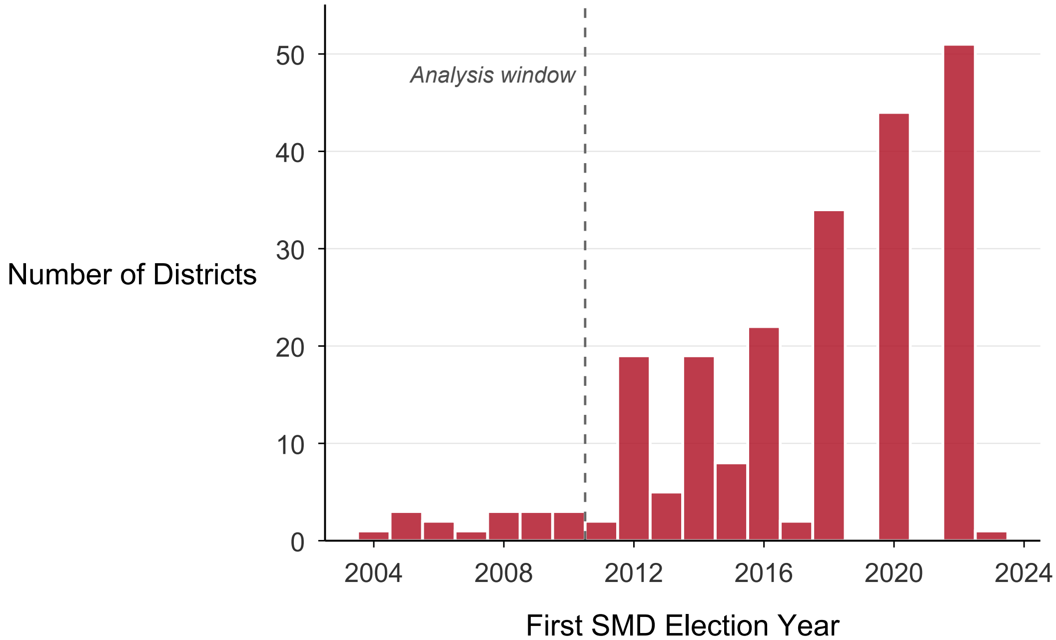


Figure 1: Treatment Timing: Staggered Adoption of Single-Member District Elections
Notes: Distribution of 230 treated school districts by year of first SMD election. The eight core CVRA-era cohorts (2012–2020) contain 156 districts. Twenty-two early adopters (1999–2010) predate the analysis window; 52 late adopters (2022–2023) have limited post-treatment data.

representation is hypothesized to reduce differential treatment across racial groups (Meier and England 1984; Skiba et al. 2011). Gaps also difference out district-wide policy changes (e.g., a universal shift toward restorative practices) that affect all students equally, isolating race-specific treatment effects. I therefore pair gap estimates with race-specific rate estimates in the main results, so the analysis distinguishes relative disparity effects from absolute changes in each group's suspension risk. The Black-White gap is valid only for district-years in which both Black and White enrollment exceed a minimum threshold (10 students), yielding approximately 60% coverage of the analysis panel. The Hispanic-White gap has approximately 93% coverage, reflecting the larger Hispanic population in California. I also examine the total suspension rate as a secondary outcome.

Academic years 2019–20 and 2020–21 are excluded from all analyses due to the COVID-19 pandemic, which substantially disrupted both school operations and discipline practices. The final analysis panel contains 9,439 non-COVID district-year observations from 788 districts.

3.4 Identification Strategy

Identification relies on a staggered difference-in-differences design. The baseline specification is:

$$Y_{dt} = \alpha_d + \lambda_t + \beta \cdot \mathrm{SMD}_{dt} + \varepsilon_{dt} \tag{1}$$

where Y_{dt} is the discipline outcome for district d in year t, α_d and λ_t are district and year fixed effects, SMD_{dt} is an indicator equal to one for treated districts in post-treatment years, and ε_{dt} is the disturbance. The coefficient β captures the average treatment effect on the treated (ATT). Standard errors are clustered at the district level throughout.

Identification requires that, conditional on district and year fixed effects, treated and control districts would have followed parallel trends in the absence of treatment. I assess this assumption through event study specifications:

$$Y_{dt} = \alpha_d + \lambda_t + \sum_{k \neq -1} \gamma_k \cdot \mathbf{1}[\mathrm{rel\_time}_{dt} = k] + \varepsilon_{dt} \tag{2}$$

where $\mathrm{rel\_time}_{dt}$ is the number of years relative to treatment ($k = 0$ is the first treatment year), and the omitted category is $k = -1$. Under parallel trends, the pre-treatment coefficients γ_k for $k < -1$ should each be zero. I report event study estimates from the heterogeneity-robust estimators described below; each modifies the aggregation of cohort-specific relative-time effects to avoid the negative-weighting problem inherent in the TWFE version of Equation 2.

The exogeneity of treatment timing is the key identifying assumption. CVRA litigation timing is driven by plaintiff attorneys' capacity constraints and case-selection strategies, not by changes in districts' discipline policies. While the *pool* of eligible districts is selected on demographics (size, diversity, segregation), the *sequence* of challenges within this pool is plausibly orthogonal to discipline trends. The identifying assumption is one of parallel *trends*, not parallel *levels*: two groups can have identical levels but diverging trajectories. I therefore assess this assumption primarily by examining pre-treatment dynamics in the event study (Section 4). Table 1 additionally shows that treated and control districts have broadly similar pre-treatment discipline levels, though I note that similar levels are necessary but not sufficient for parallel trends, and that the normalized differences are non-trivial on enrollment and on the BW gap itself.

3.5 Estimators

Recent econometric work has demonstrated that the TWFE estimator in Equation 1 can be biased when treatment effects are heterogeneous across cohorts or over time (Goodman-Bacon 2021; de Chaisemartin and D'Haultfœuille 2020). The bias arises because TWFE implicitly uses already-treated units as comparisons for newly-treated units, potentially creating "negative weights" on some treatment effects.

I therefore supplement the TWFE estimates with five alternative estimators. The *Sun-Abraham* (SA) estimator uses interaction-weighted averages of cohort-specific treatment effects, restricting comparisons to never-treated and not-yet-treated units (Sun and Abraham 2021). I designate this as the preferred estimator. The *Callaway-Sant'Anna* (CS) estimator constructs group-time average treatment effects for each cohort-period pair, then aggregates to an overall ATT (Callaway and Sant'Anna 2021). The *Borusyak-Jaravel-Spiess* (BJS) imputation estimator first estimates the counterfactual outcomes for treated observations using only untreated data, then takes the average difference between observed and imputed outcomes as the ATT (Borusyak, Jaravel, and Spiess 2024). I also implement Synthetic Difference-in-Differences (SDID), which combines the unit-weight approach of synthetic control with the time-differencing of DiD (Arkhangelsky et al. 2021). Finally, I report estimates from the Generalized Synthetic Control (GSC) method, which models untreated potential outcomes using interactive fixed effects (Xu 2017). SDID and GSC serve as robustness checks under alternative identifying assumptions; because they absorb interactive (unit-specific) trends, they may over-correct if the true data-generating process does not contain such trends.

All standard errors are clustered at the district level. I apply Benjamini-Yekutieli corrections for multiple testing across the three primary outcomes (Benjamini and Yekutieli 2001).

4 Results

4.1 Descriptive Statistics and Balance

Table 1 presents summary statistics for the analysis sample. Treated districts are substantially larger than control districts (mean enrollment 11,271 vs. 3,437; normalized difference = 1.01), reflecting the CVRA's focus on larger, more diverse jurisdictions. Treated districts also have somewhat larger Black-White suspension gaps (9.76 vs. 7.13 percentage points; normalized difference = 0.21) and Hispanic-White gaps (1.01 vs. -0.40; normalized difference = 0.27). Total suspension rates are modestly higher in treated districts (6.95 vs. 5.57; normalized difference = 0.20). White suspension rates are similar across groups (6.15 vs. 5.83; normalized difference = 0.04), while race-specific rates for Black students (15.96 vs. 13.28; normalized difference = 0.16) and Hispanic students (7.15 vs. 5.43; normalized difference = 0.26) are elevated in treated districts.

The normalized differences, while non-trivial for enrollment, are generally moderate for discipline outcomes, suggesting that the two groups are reasonably comparable conditional on district and year fixed effects. The enrollment imbalance is a mechanical consequence of the CVRA targeting large, diverse districts and does not threaten identification so long as treatment timing within the eligible pool is exogenous, which I argue above.

Figure 2 displays the raw Black-White suspension gap over time for treated and control districts. Before the modal treatment window (2016–2020, shaded), the two groups track each other closely, with treated districts maintaining a consistently higher BW gap

Table 1: Summary Statistics and Balance

Variable	Full Sample Mean	Full Sample SD	By Treatment Status Control	By Treatment Status Treated	Norm. Diff.
Enrollment (total)	5,583	9,021	3,437	11,271	1.01
Total suspension rate	5.86	6.78	5.57	6.95	0.20
Black suspension rate	14.16	15.98	13.28	15.96	0.16
Hispanic suspension rate	5.80	6.39	5.43	7.15	0.26
White suspension rate	5.91	7.40	5.83	6.15	0.04
BW suspension gap (pp)	8.09	12.38	7.13	9.76	0.21
HW suspension gap (pp)	−0.10	5.23	−0.40	1.01	0.27
Sample sizes					
Districts	788		558	230	
District-year obs (non-COVID)	9,439				

Notes: Suspension rates are suspensions per 100 enrolled students. BW gap = Black rate − White rate. HW gap = Hispanic rate − White rate. Normalized difference = (Treated mean − Control mean) / $\sqrt{(s_T^2 + s_C^2)/2}$. Sample restricted to non-COVID years (excluding 2019–20 and 2020–21). BW gap available for ≈60% of observations due to minimum Black enrollment requirements.

of approximately 2–3 percentage points. Beginning around 2016, the groups diverge: the treated-group BW gap rises sharply while the control-group gap declines gradually. By 2024, the gap between groups exceeds 5 percentage points. This visual parallel in pre-treatment trends motivates the DiD design, while the divergence is consistent with a treatment effect. For comparison, Figure A13 in Appendix B shows race-specific suspension rates by treatment group.

4.2 Main Effects

Table 2 presents the main difference-in-differences estimates. I focus first on the Black-White suspension gap (column 1). The TWFE estimate is +1.28 percentage points (SE = 0.67; $p = .058$), marginally significant. The preferred Sun-Abraham estimate is +2.74 percentage points (SE = 0.92; $p = .003$). Callaway-Sant'Anna yields +2.17 (SE = 0.87; $p = .013$). BJS imputation yields a smaller, insignificant +1.20 (SE = 0.74; $p = .102$), likely reflecting sensitivity to the imputation model in this many-cohort setting. Synthetic DiD produces

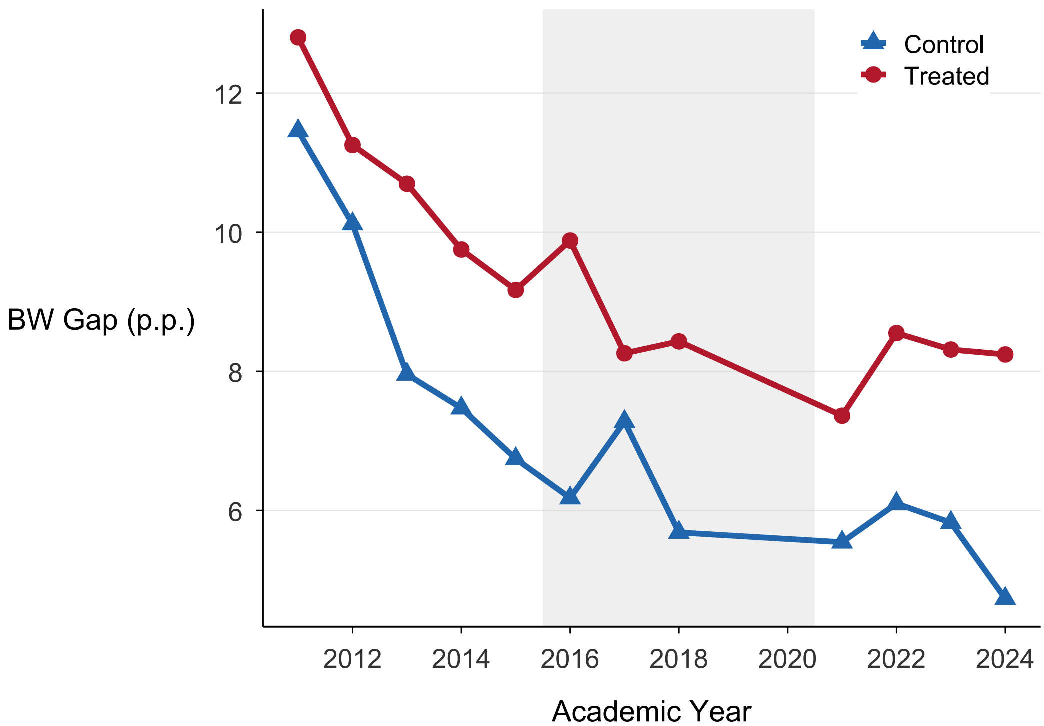


Figure 2: Raw Trends: Black-White Suspension Rate Gap
Notes: Annual mean Black-White suspension rate gap (percentage points) for treated and control districts. The shaded band marks the modal treatment window (2016–2020). COVID years (2019–20, 2020–21) are excluded.

a similarly attenuated estimate of +0.61 (SE = 0.64). The Benjamini-Yekutieli adjusted q-value for the BW gap is 0.023, confirming significance after multiple-testing correction.

All point estimates are positive, indicating that the Black-White suspension gap increases after SMD adoption. The TWFE estimate is attenuated relative to the heterogeneity-robust estimators, consistent with the negative weighting bias that arises when already-treated units serve as controls for late adopters (Goodman-Bacon 2021).

The Hispanic-White gap (column 2) shows no robust effect. TWFE yields -0.46 (SE = 0.23; $p < .05$), but all heterogeneity-robust estimators are insignificant: SA = -0.16 (SE = 0.47), CS = -0.27 (SE = 0.50), BJS = -0.37 (SE = 0.29), SDID = -0.18 (SE = 0.23). The BY-adjusted q-value is 0.963. The reversal—TWFE significant but robust estimators null—is the mirror image of the BW gap pattern, and likely reflects Goodman-Bacon contamination operating in the opposite direction: for the HW gap, the problematic already-treated-vs.-later-treated comparisons happen to amplify the estimate rather than attenuate it.[2]

The total suspension rate (column 3) is likewise unaffected: all estimates are small and insignificant (SA = −0.15, SE = 0.34). To calibrate these nulls, minimum detectable effects at 80% power (based on SA standard errors) are 1.3 pp for the HW gap and 1.0 pp for the total rate—23% and 15% of the respective treated-group means. The HW and total rate nulls are therefore informative, not simply underpowered.

Race-specific rate regressions show that the BW gap increase is driven by higher Black suspension rates (+1.80 pp under TWFE, $p = .037$; SA estimates are directionally similar), while White (+0.10; n.s.) and Hispanic (−0.33; n.s.) rates are near zero. This implies the average treated district sees a rise in Black students' absolute suspension risk, not only a relative gap shift.[3] This asymmetry—effects on Black but not Hispanic or White outcomes—anticipates the mechanism analysis in Section 5: of the 230 treated districts, only 17 gained a predicted-Black board member after SMD adoption, while 163 gained a predicted-Hispanic member. The average effect is thus dominated by districts where the newly elected representatives do not share the racial background of the most affected student group.

Figure 3 provides an at-a-glance comparison of all five estimators across the three primary outcomes, illustrating both the consistency of the BW gap finding and the clean nulls for the HW gap and total rate.

Interactive fixed effects methods. The SDID estimate for the BW gap (+0.61; SE = 0.64) and the Generalized Synthetic Control estimate (+0.80; SE = 0.93; $p = .39$; nonparametric bootstrap with 1,000 replications) are smaller and statistically insignificant, suggesting that some portion of the DiD effect may reflect differential trends that these methods absorb (see

[2]This is plausible because the sign and magnitude of TWFE bias depend on the covariance between treatment timing and treatment-effect dynamics, which can differ across outcomes (Goodman-Bacon 2021).

[3]The race-specific rate regressions are estimated on the full analysis panel, while the BW gap regression is restricted to district-years with at least 10 Black *and* 10 White students. The two samples are therefore not directly comparable, and the race-specific rates should not be read as an exact arithmetic decomposition of the BW gap estimate.

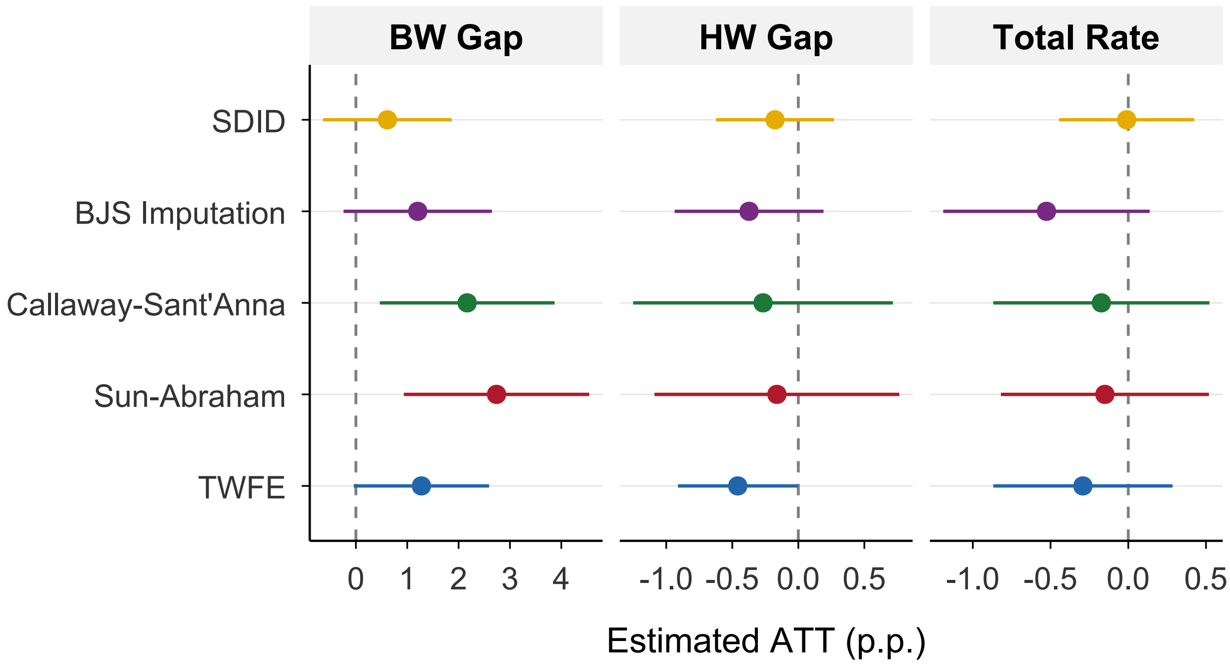


Figure 3: ATT Estimates Across Estimators and Outcomes
Notes: Point estimates with 95% confidence intervals for each estimator-outcome combination. TWFE = two-way fixed effects; SA = Sun-Abraham; CS = Callaway-Sant'Anna; BJS = Borusyak-Jaravel-Spiess imputation; SDID = Synthetic DiD. All specifications include district and year fixed effects and exclude COVID years.

Appendix C for SDID trajectories and diagnostics). Cross-validation selected zero latent factors for the BW and HW gap outcomes ($r^* = 0$) and one factor for the total rate ($r^* = 1$). The $r^* = 0$ result indicates that the data do not support interactive fixed effects beyond standard unit and time effects for the primary outcomes. SA remains the preferred estimator for reasons detailed in Section 4.4. Because SDID and GSC absorb interactive trends by construction, they may over-correct in settings where the true data-generating process does not contain such trends—and the $r^* = 0$ result is consistent with this concern. The IFE estimates are therefore best interpreted as a lower bound on the effect under an alternative, more conservative set of identifying assumptions.[4]

[4]GSC removes treated units with fewer than $r^* + 2$ pre-treatment periods, yielding effective samples of 311 (BW gap), 548 (HW gap), and 611 (total rate) of the 377–700 balanced-panel units. All removed units are treated; control units are observed in all years. Under GSC, the HW gap shows a marginally significant negative estimate (-0.65; SE = 0.25; $p = .010$ nonparametric), suggesting a possible reduction in the Hispanic-White gap that the DiD-family estimators—which retain all treated units—do not detect. This result should be interpreted cautiously given the substantial sample attrition.

Table 2: Main Difference-in-Differences Results

	BW Gap (1)	HW Gap (2)	Total Rate (3)
Panel A: Estimators			
TWFE	1.28*	−0.46**	−0.29
	(0.67)	(0.23)	(0.29)
Sun-Abraham	2.74***	−0.16	−0.15
	(0.92)	(0.47)	(0.34)
Callaway-Sant'Anna	2.17**	−0.27	−0.17
	(0.87)	(0.50)	(0.35)
BJS Imputation	1.20	−0.37	−0.53
	(0.74)	(0.29)	(0.34)
Synthetic DiD	0.61	−0.18	−0.01
	(0.64)	(0.23)	(0.22)
Panel B: Inference			
BY q-value (SA)	0.023	0.963	1.000
N (obs)	≈5,632	≈8,765	≈9,325

Notes: Each cell reports the ATT estimate from the indicated estimator with standard errors clustered at the district level in parentheses. BW Gap = Black-White suspension rate gap (pp). HW Gap = Hispanic-White suspension rate gap (pp). Total Rate = overall suspension rate per 100 students. All specifications include district and year fixed effects and exclude COVID years (2019–20, 2020–21). BY q-values are Benjamini-Yekutieli adjusted across three primary outcomes. $^{*}p < .10$; $^{**}p < .05$; $^{***}p < .01$.

4.3 Event Study

Figure 4 presents the Sun-Abraham event study for the Black-White suspension gap over a $[-5, +5]$ window. The pre-treatment coefficients are small and non-monotonic: $t{-}5 = -0.53$, $t{-}4 = +1.29$, $t{-}3 = -0.34$, $t{-}2 = +0.68$ (relative to $t{-}1 = 0$). None is individually significant, and the oscillating pattern is inconsistent with a systematic pre-trend.

Post-treatment, the effect emerges at $t = 0$ (+1.28) and increases over time: $t{+}1 = +1.79$, $t{+}2 = +2.34$ ($p < .10$), $t{+}3 = +3.32$ ($p < .01$), $t{+}4 = +5.02$ ($p < .01$), $t{+}5 = +2.67$ ($p < .05$). The gradual ramp-up is consistent with a mechanism that operates through changes in board composition and subsequent policy adjustments, both of which require time to materialize. The peak at t+4 followed by partial attenuation at t+5 may reflect the limited number of districts observed at longer horizons.

Figure 5 presents event studies for all primary outcomes, comparing SA, CS, and BJS estimators. The three modern estimators produce similar event study patterns for the BW gap, lending confidence that the results are not driven by a particular estimator's assumptions. The HW gap and total rate event studies show flat dynamics, consistent with the null average effects in Table 2.

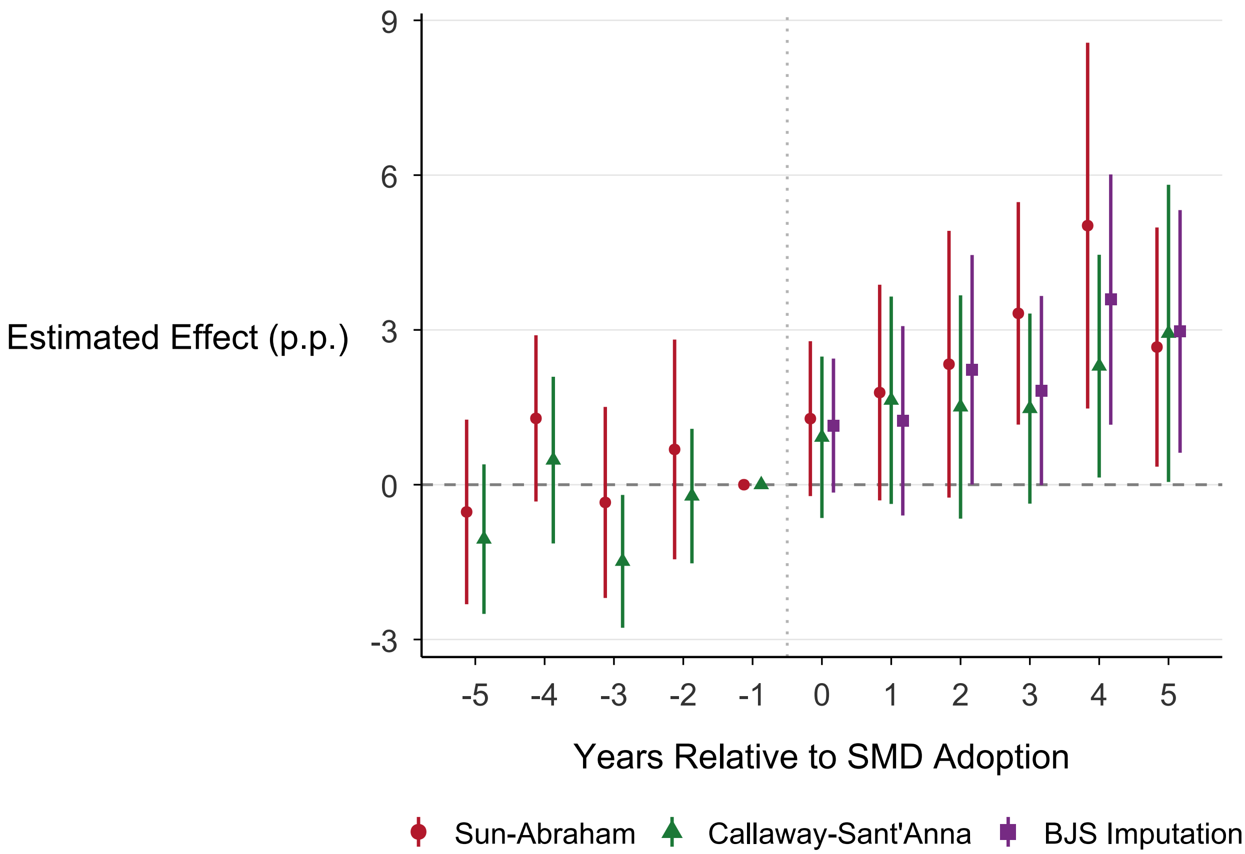


Figure 4: Event Study: Effect of SMD Adoption on Black-White Suspension Gap
Notes: Coefficients from Sun-Abraham, Callaway-Sant'Anna, and BJS imputation event study specifications. The dependent variable is the Black-White suspension rate gap in percentage points. The omitted period is $t-1$. Whiskers show 95% confidence intervals based on district-clustered standard errors. COVID years (2019–20, 2020–21) are excluded.

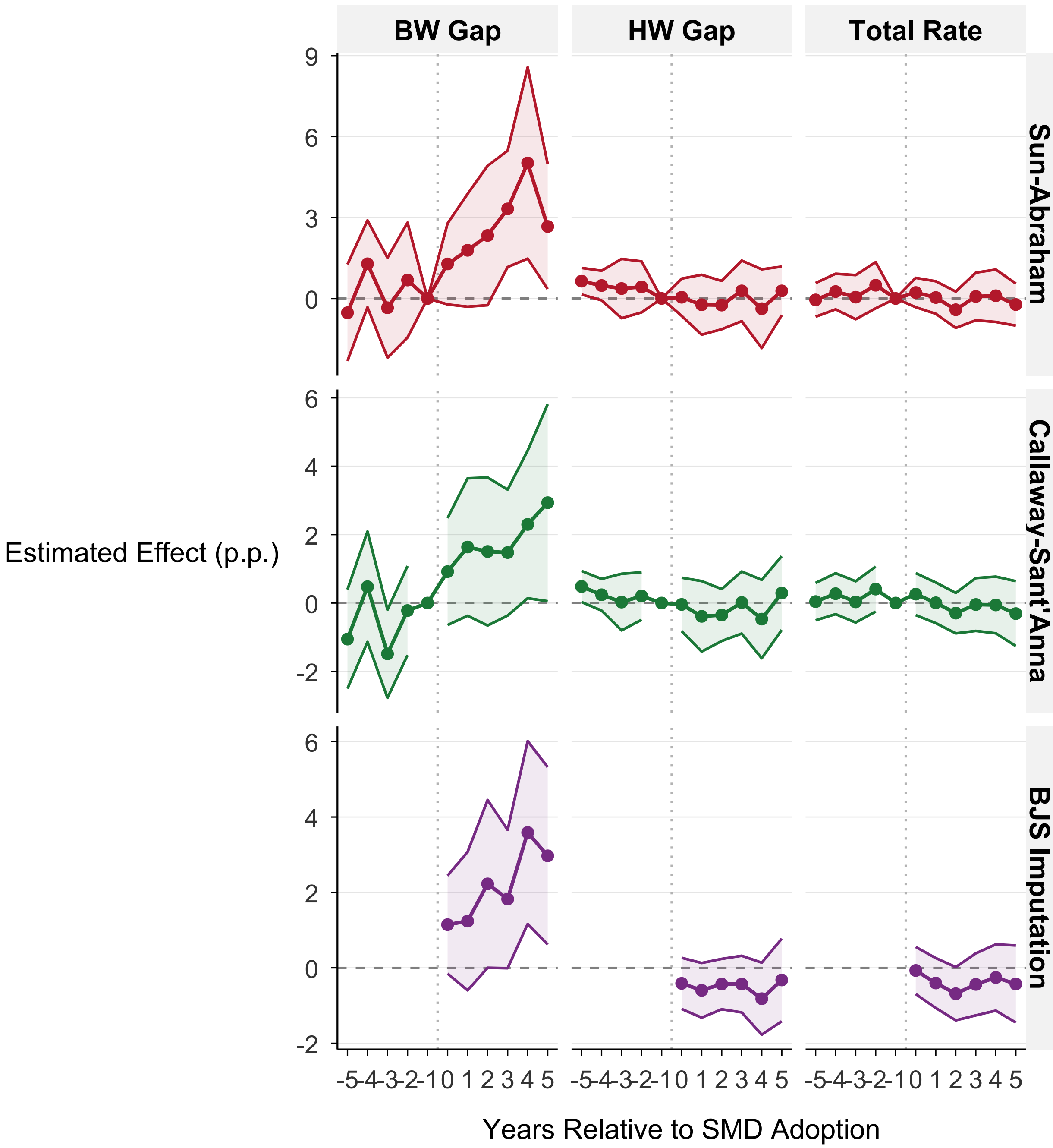


Figure 5: Event Studies: All Primary Outcomes

Notes: Event study coefficients for BW gap, HW gap, and total suspension rate across three heterogeneity-robust estimators (Sun-Abraham, Callaway-Sant'Anna, BJS Imputation). Shaded bands show 95% confidence intervals. The omitted period is $t-1$. The BW gap shows a clear post-treatment ramp-up across all estimators; the HW gap and total rate are flat, consistent with the null average effects in Table 2. COVID years excluded.

4.4 Robustness

I conduct an extensive battery of robustness checks, summarized in Figure 7 and detailed in Appendix D. I focus here on the tests most relevant to identification.

Design-based inference. Randomization inference—reassigning treatment timing 500 times—yields $p = .004$ (2 of 500 permutations exceed the observed ATT), confirming significance under the sharp null without parametric assumptions. A triple-difference specification that exploits within-district, across-race variation to difference out district-level trends produces an even larger estimate (+3.12; $p = .003$).

Specification stability. Across nine alternative specifications (Figure 7)—including pre-COVID-only, excluding large districts, higher enrollment thresholds, clean switchers, balanced panel, district-specific trends, and outlier trimming—the SA BW gap ranges from +2.3 to +3.9 pp, always significant. The estimate is stable to controlling for poverty (FRPM percentage), English Learner shares, ACS income/poverty, per-pupil spending, and interacted year fixed effects; see Appendix D.

Sensitivity to Black enrollment thresholds. The baseline BW gap estimate requires at least 10 Black and 10 White students per district-year. Raising this threshold attenuates the estimate: the SA ATT declines from +2.86 (SE = 0.81; $p < .001$) at the baseline to +2.52 ($p = .001$) at ≥ 20, +1.74 ($p = .024$) at ≥ 50, and +1.15 ($p = .13$) at ≥ 100 Black students. The effect remains significant at the ≥ 50 threshold but not at ≥ 100, where only 284 districts (149 treated) contribute. The attenuation pattern is consistent with rate instability in districts with small Black populations inflating the baseline estimate. The policy-relevant effect size in districts with substantial Black enrollment (≥ 50 students) is approximately +1.7 pp, roughly two-thirds of the headline estimate.

Placebo and falsification tests. Four placebo outcomes (total enrollment, White suspension rate, Asian suspension rate, percent Black enrollment) are all null, ruling out generic post-treatment divergence. Using SEDA 6.0 achievement gap data (Reardon et al. 2026), standardized test score gaps are similarly null (W-H: +0.012 SD, $p = .33$; W-B: +0.004

SD, $p = .78$), while the discipline BW gap remains significant on the same sample (+3.29 pp; $p = .01$; Appendix H). The Benjamini-Yekutieli adjusted q-value for the BW gap is 0.023.

Cohort stability. Leave-one-cohort-out analysis (Figure 6) shows the SA BW gap ranging from +1.7 (dropping 2012) to +3.4 (dropping 2016), positive and statistically significant in every iteration. Interactive fixed effects methods (SDID: +0.61; GSC: +0.80, SE = 0.93, $p = .39$) yield smaller, insignificant BW gap estimates, suggesting some role for differential trends; three considerations favor the DiD estimates (non-monotonic pre-trends, larger triple-diff, RI significance), and the IFE estimates serve as conservative bounds (Appendix C).

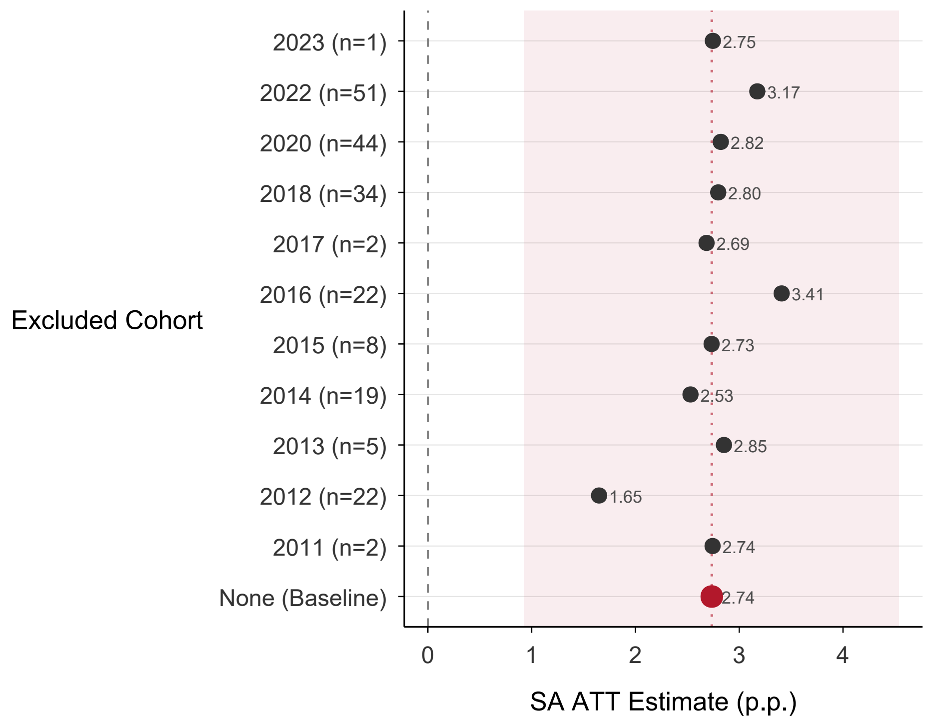


Figure 6: Leave-One-Cohort-Out: BW Suspension Gap (Sun-Abraham)
Notes: SA ATT estimates for the BW suspension gap when each treatment cohort is excluded in turn. The "None" row shows the baseline estimate including all cohorts. The vertical dashed line marks the baseline estimate. Cohort sizes reflect the BW-gap-valid subsample and may differ slightly from the full-sample counts reported in Section 3.

Pre-trends and sensitivity. I address the parallel trends assumption with honesty about its limitations. While individual pre-treatment event study coefficients are small and non-monotonic, a joint F-test on all pre-treatment SA coefficients rejects the null of zero

($p < .001$). This rejection, however, likely reflects the high power of the test combined with small but nonzero pre-treatment coefficients rather than a substantively meaningful violation.

To formalize this concern, I implement the Rambachan and Roth (2023) HonestDiD sensitivity analysis, which constructs robust confidence intervals under controlled violations of parallel trends. Figure 8 presents the results under smoothness restrictions (ΔSD), which bound the maximum change in the trend violation between consecutive periods. The robust confidence interval for the BW gap crosses zero at a smoothness parameter of $M = 0.1$, meaning that if the maximum post-treatment change in the trend violation exceeds 10% of the maximum pre-treatment change, the confidence interval includes zero. Under the complementary relative magnitudes approach (ΔRM), the interval includes zero at all $\bar{M} \geq 0.5$ (Appendix Figure A22).

This sensitivity analysis reveals a genuine limitation of the design. I note, however, several mitigating factors. First, the individual pre-treatment coefficients are small in absolute magnitude and non-monotonic, suggesting noise rather than systematic trending. Second, randomization inference—which provides design-based inference under exchangeability of treatment timing—yields $p = .004$. Third, the triple-difference specification eliminates district-level trends that are common across racial groups and produces an even larger estimate (+3.12); however, it does not eliminate race-specific district trends, which remain a potential confound. Fourth, the mechanism evidence presented in Section 5 is suggestive of a causal interpretation—the finding that BW gaps narrow specifically where Black representatives are elected is consistent with a causal channel from representation to outcomes—though the moderation analysis relies on the same parallel trends assumption and is not independent of the main identification strategy. I return to this limitation in the conclusion.

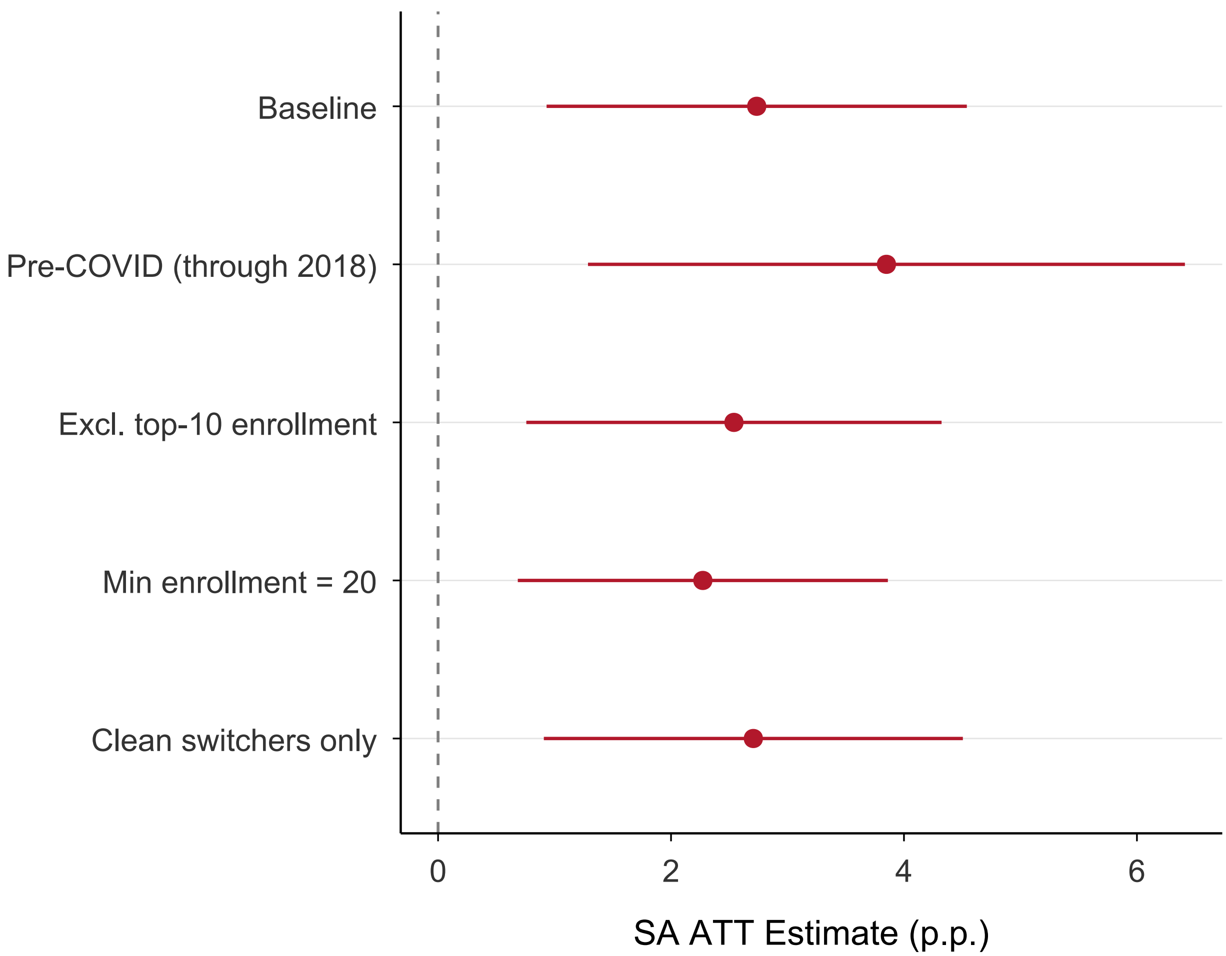


Figure 7: Robustness: Sun-Abraham ATT Estimates Across Specifications
Notes: Each row shows the Sun-Abraham ATT estimate for the BW suspension gap under a different specification. Points are estimates; whiskers show 95% confidence intervals. The vertical dashed line marks zero. All estimates are positive and statistically significant.

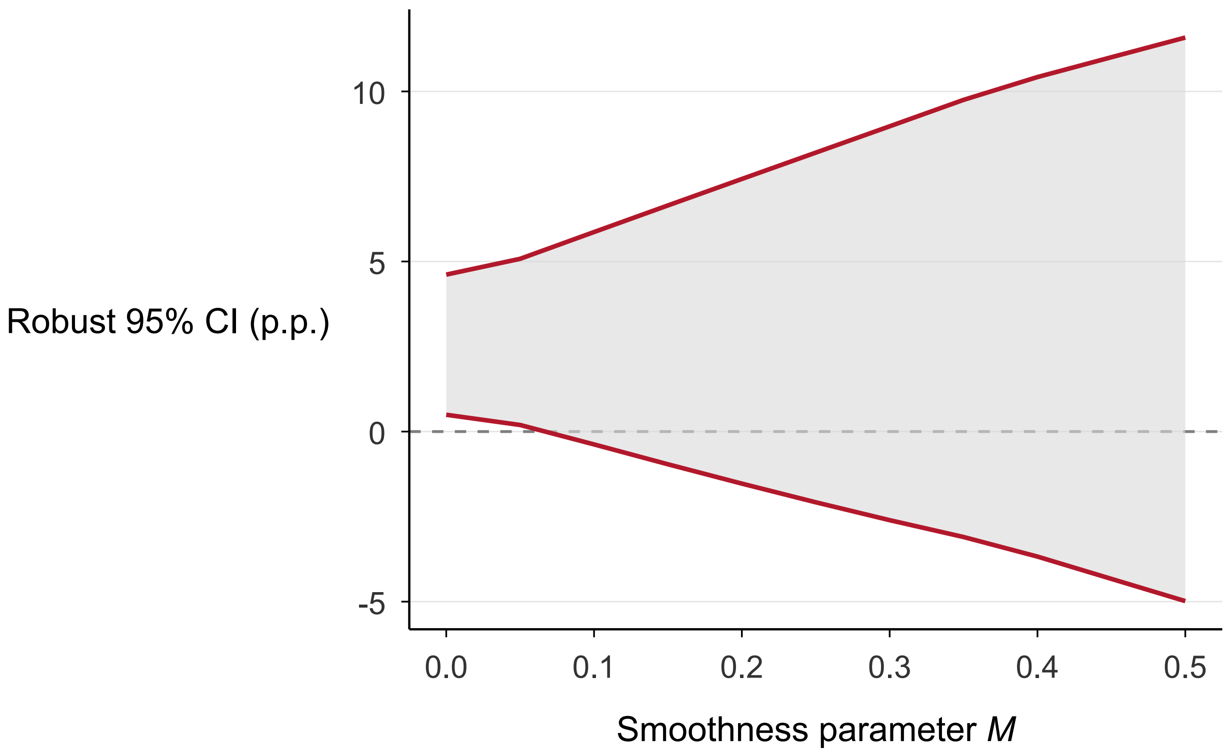


Figure 8: Sensitivity to Parallel Trends Violations (HonestDiD)

Notes: Robust confidence sets from Rambachan and Roth (2023) for the BW suspension gap effect under smoothness restrictions on the degree of parallel trends violations. The x-axis shows the smoothness parameter M, which bounds the maximum change in the trend violation between consecutive periods relative to the maximum observed pre-trend. At $M = 0$, the standard parallel trends assumption holds; the robust CI crosses zero at approximately $M = 0.1$.

5 Mechanisms: The Race-Match Channel

This section investigates the mechanism behind the divergent effects on Black-White and Hispanic-White gaps by examining three links in the causal chain: (1) the "first stage" effect of SMD elections on board composition, (2) the moderating role of board racial composition on discipline outcomes, and (3) the parallel channel of teacher workforce diversity.

5.1 First Stage: SMD Elections and Board Composition

Table 3 documents the effect of SMD adoption on the racial composition of elected school board members. Panel A reports DiD estimates. The TWFE estimate shows that SMD adoption increases the overall minority (non-White) share of elected board members by 3.4 percentage points (SE = 1.8; $p = .055$). Decomposing by group, the effect is concentrated among Hispanic candidates: the Hispanic share increases by 4.1 percentage points (SE = 1.9; $p = .035$). The Black share is essentially unchanged (+0.2 pp; n.s.). Sun-Abraham estimates are directionally consistent but insignificant (minority: +2.6 pp, SE = 2.5; Hispanic: +3.8 pp, SE = 3.0; Black: −0.4 pp, SE = 0.8), reflecting the smaller effective sample available after restricting comparisons to never-treated units. These estimates are consistent with Abott and Magazinnik (2020), who find significant effects on Latino representation but not Black representation.

Panel B provides a complementary view using raw pre-post comparisons. Among elected board members in treated districts, the BISG-predicted minority probability averages 47.1% before SMD adoption and 56.7% after—an increase of 9.6 percentage points. The Hispanic share increases from 26.7% to 36.7% (+10.0 pp), while the Black share declines slightly from 6.7% to 6.0% (−0.6 pp). These patterns are consistent across 223 districts with pre-SMD elections and 222 with post-SMD elections, based on 3,522 pre-SMD and 1,356 post-SMD elected candidates. The gap between the raw pre-post difference (+9.6 pp) and the DiD estimate (+3.4 pp) implies that control districts also saw roughly a 6 pp increase in

Table 3: First Stage: Electoral Reform and Board Composition

	Minority Share	Hispanic Share	Black Share
Panel A: DiD Estimates			
TWFE	0.034*	0.041**	0.002
	(0.018)	(0.019)	(0.006)
Sun-Abraham	0.026	0.038	−0.004
	(0.025)	(0.030)	(0.008)
Panel B: Pre-Post Means			
Pre-SMD	0.471	0.267	0.067
Post-SMD	0.567	0.367	0.060
Difference	+0.096	+0.100	−0.006
N candidates (Pre / Post)		3,522 / 1,356	
N districts (Pre / Post)		223 / 222	

Notes: Panel A reports DiD estimates of the effect of SMD adoption on the BISG-predicted racial composition of elected board members. Standard errors clustered at the district level in parentheses. SA estimates are insignificant for all three outcomes; I retain TWFE as the primary specification for the first stage since the discrete, unambiguously positive nature of the compositional shift makes it less susceptible to the negative-weighting concerns that affect the discipline outcomes. Panel B reports raw mean BISG-predicted race probabilities for elected candidates before and after SMD adoption. $^{*}p < .10$; $^{**}p < .05$; $^{***}p < .01$.

minority board share over the same period, driven by statewide demographic and political trends. The CVRA's causal contribution to representational change is thus approximately one-third of the raw compositional shift.

Figure 9 visualizes the dynamics of board composition around SMD adoption. The minority and Hispanic shares show clear increases beginning at $t = 0$ and persisting through $t+5$, while the Black share remains flat. This pattern confirms that the CVRA reform wave in California is primarily a story about Hispanic empowerment, not minority empowerment broadly.

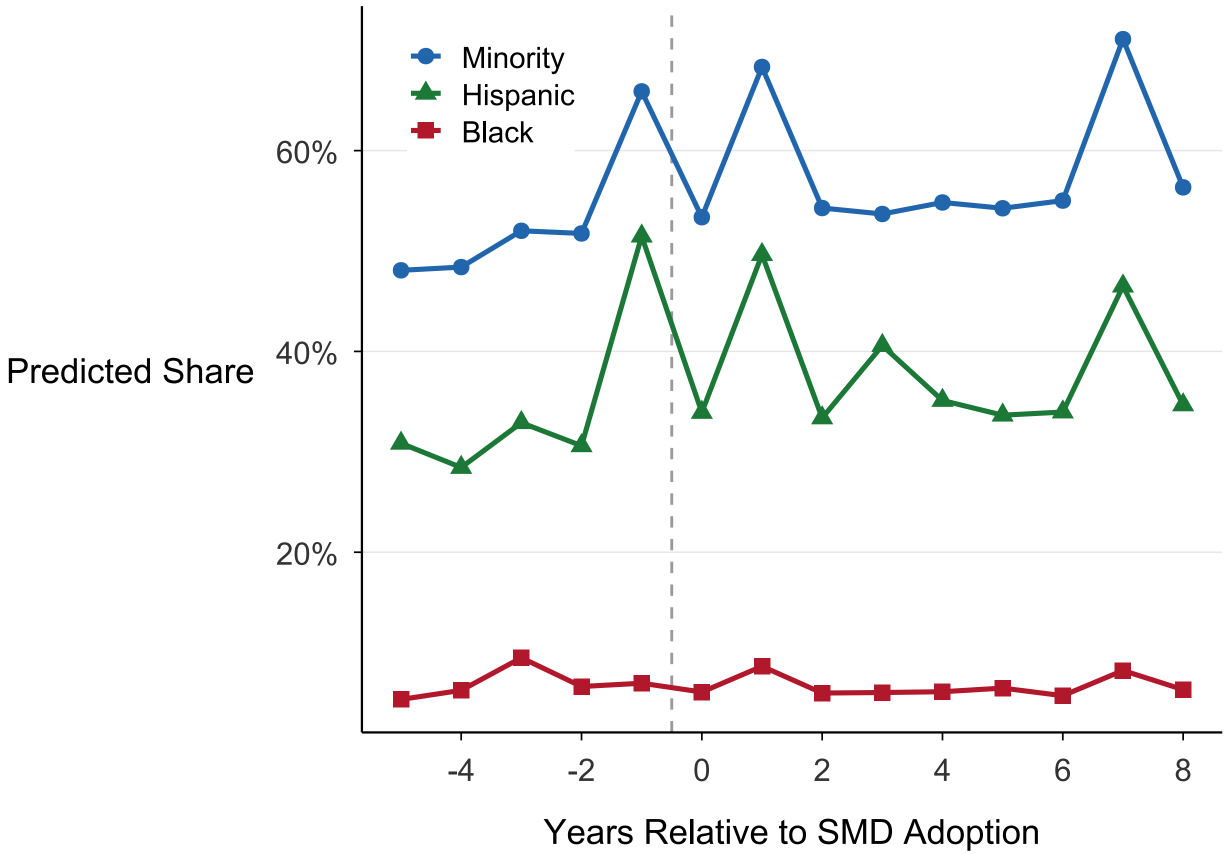


Figure 9: Board Composition Dynamics Around SMD Adoption
Notes: Mean BISG-predicted racial shares of elected board members by event time relative to SMD adoption. The dashed vertical line marks the first SMD election. Minority and Hispanic shares increase sharply at adoption; Black shares remain flat.

5.2 Board Composition and Discipline Outcomes

If the race-match condition holds, then the BW gap should decline in districts that gain Black board members and be unaffected in districts that gain only Hispanic members. I test this prediction using moderation models that interact the treatment indicator with measures of board racial composition.

Table 4, Panel A, reports the results. The key specification interacts treat_post with the cumulative Black share of the school board (i.e., the running mean of BISG-predicted Black probability among elected members from the first SMD election onward). The base ATT—the effect at zero Black board share—is +3.56 ($p < .001$), indicating a significant increase in the BW gap when no Black member is on the board. The interaction with Black board share is -33.7 (SE $= 11.9$; $p = .005$), indicating that each percentage-point increase in Black board representation is associated with a 0.34 pp reduction in the BW gap. At the highest observed levels of Black board share, the interaction is large enough to offset the base ATT, though most districts with any Black representation gained too little to fully eliminate the gap increase (see reconciliation below).

The mismatch test—interacting the SMD treatment indicator with Hispanic board share, with the BW gap as outcome—yields a null result (-0.6; $p = .83$), confirming that Hispanic representation does not reduce Black-White disparities. Analogously, the HW gap is not significantly moderated by Hispanic board share (interaction $= -1.2$; $p = .38$) or by Black board share ($+1.6$; $p = .69$). The total suspension rate shows a marginally significant moderation by overall minority board share (-1.9; SE $= 1.1$; $p = .08$), suggesting that minority representation may reduce exclusionary discipline broadly. The board composition channel thus operates most strongly through Black representation on Black-White outcomes—consistent with a race-match mechanism—but minority board share also exerts a general moderating influence on total suspension rates.

Figure 10 visualizes this race-match moderation by plotting the marginal effect of SMD adoption on the BW gap as a function of cumulative Black board share. At zero Black representation, the treatment effect is strongly positive (+3.6 pp). As Black board share increases, the effect declines linearly, crossing zero at approximately 11% Black share. The shaded 95% confidence band excludes zero for low levels of Black representation but widens at higher levels where data are sparse (only 17 treated districts gained a Black board member).

The heterogeneity analysis further supports this pattern. Of the 230 treated districts, only 17 gained a predicted-Black board member after SMD adoption, compared to 140 that gained a predicted-Hispanic but not Black member. The SA estimate for districts that gained a Black member is $+2.16$ ($p < .01$), while for districts that gained a Hispanic but not Black member, the estimate is $+3.10$ ($p < .01$; see Figure A32). The difference between these

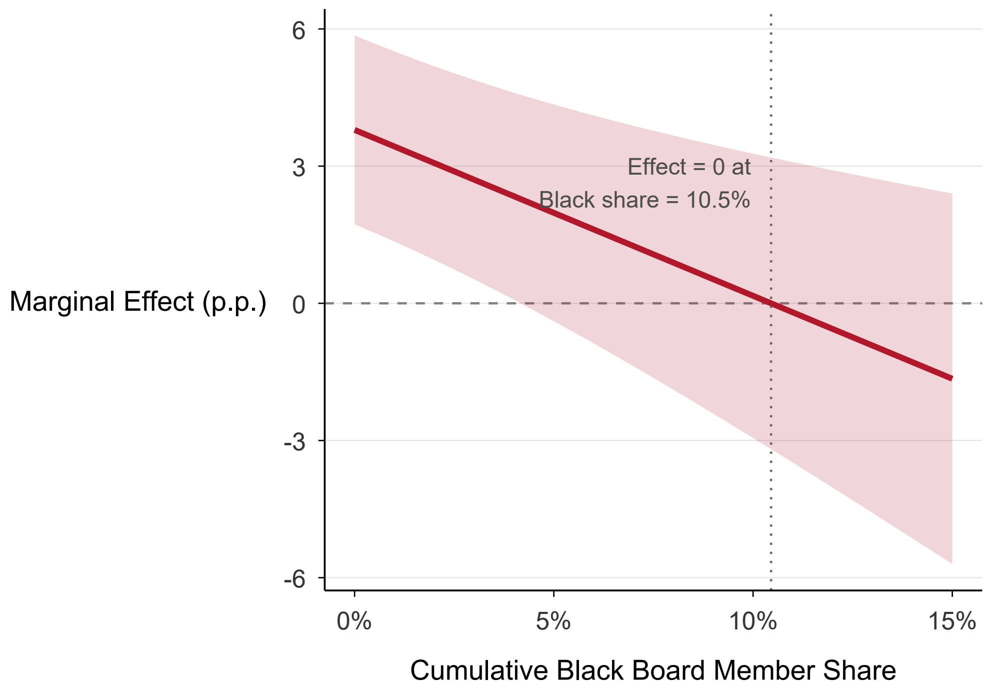


Figure 10: Race-Match Moderation: Marginal Effect of SMD on BW Gap by Black Board Share
Notes: Marginal effect of SMD adoption on the Black-White suspension gap (p.p.) as a function of cumulative Black board member share, from the moderation model in Table 4. The shaded band shows 95% confidence intervals. The dotted vertical line marks the Black share at which the marginal effect equals zero.

subgroup estimates—and the larger effect in the Hispanic-only group—is consistent with the race-match theory: without Black-specific advocacy, the BW gap widens more.

It is important to reconcile the subgroup and interaction estimates. The positive subgroup ATT for "gained Black member" districts (+2.16) does not contradict the large negative interaction coefficient (−33.7). The subgroup estimate is the average effect across districts with any Black representation gain, but these districts vary widely in how much Black representation they gained. Most gained only one Black member on a five- or seven-member board, yielding a cumulative BISG-probability Black share (the running mean of posterior Black probabilities across all post-SMD elected members) far below the approximately 10% at which the interaction model predicts the gap would narrow to zero. The interaction model thus indicates that Black representation attenuates the BW gap increase—reducing it from +3.6 pp at zero Black share to approximately +2.2 pp at the average Black share among these 17 districts—but does not fully offset it in most cases. In districts with sustained or substantial Black representation, the moderation effect is large enough to eliminate or reverse the gap increase, but such districts are rare in the California sample.

5.3 Teacher Workforce Diversity

The board composition channel may operate in part through hiring decisions that diversify the teacher workforce. I examine this pathway using CDE staff demographic data.

SMD adoption increases the minority teacher share by 1.5 percentage points (Sun-Abraham; $p = .029$), with the effect concentrated among Hispanic teachers (+1.4 pp; $p = .001$). Under SA, the Black teacher share shows a small, insignificant increase (+0.1 pp; SE = 0.1; $p = .22$); TWFE yields a larger, significant estimate (+0.3 pp; $p = .003$), though this may reflect the same negative-weighting bias documented for the main outcomes.[5] The student-teacher Black mismatch gap (Black student share minus Black teacher share, so that positive values indicate underrepresentation of Black teachers) declines by 0.5 percentage points under TWFE ($p = .002$), indicating improved alignment of the workforce with the student body.

Table 4, Panel B, reports the teacher diversity moderation results. The BW gap is substantially attenuated in districts with higher Black teacher shares. The interaction of treat_post with Black teacher share is -82.0 (SE = 22.3; $p = .0003$; base ATT = +3.13). With a one-year lag—allowing time for hiring decisions to take effect and new teachers to influence classroom discipline—the coefficient grows to -95.8 (SE = 24.5; $p = .0001$). These are large effects: a one-standard-deviation increase in the Black teacher share is associated with approximately a 3-percentage-point reduction in the BW gap, roughly offsetting the average treatment effect.

[5] The divergence between TWFE and SA for Black teacher share parallels the BW suspension gap pattern, where TWFE is attenuated relative to SA. For this outcome, TWFE appears to amplify rather than attenuate the estimate, consistent with the direction-ambiguity of Goodman-Bacon contamination.

The mismatch test for teachers is again null: Hispanic teacher share has no effect on the BW gap (+1.3; $p = .70$). The total suspension rate is moderated by overall minority teacher share with a one-year lag (-4.6; SE = 1.6; $p = .003$), indicating that minority teachers are associated with less exclusionary discipline overall.

Multiple testing. The mechanism analysis in this section involves 21 interaction tests across board composition, teacher diversity, three lag structures, and five outcome-moderator combinations. Applying a Benjamini-Yekutieli correction for dependent tests, 2 of 21 tests survive at $q < .05$: Black teacher share $\times$ BW gap at the 1-year lag ($q = .008$) and contemporaneously ($q = .010$). The board composition race-match interactions, while significant at raw $p < .01$, do not survive BY correction (contemporaneous $q = .094$; 1-year lag $q = .101$). Under the less conservative Benjamini-Hochberg correction, 5 of 21 tests survive, including both board composition race-match tests ($q_{BH} < .03$). The teacher diversity channel thus provides the most robust evidence for race-matched representation effects after correcting for the full family of moderator tests.

5.4 Assessing the Spending Channel

Although SMD adoption modestly increases per-pupil spending (+2.3% under SA; $p < .001$; Appendix Figure A37), I find no evidence that spending moderates the discipline gaps: the interaction with log per-pupil spending is -1.28 ($p = .41$) for the BW gap and -0.20 ($p = .73$) for the HW gap. These null interactions should be read as the absence of detectable moderation rather than a positive rejection of the spending channel—the treatment-induced spending shift of 2.3% is small, and a more powerful test would require either equivalence bounds or a larger first-stage shock. In a joint model including both spending and Black teacher share, spending is insignificant ($p = .69$) while Black teacher share remains strongly significant (-103; $p < .001$), suggesting that to the extent a discipline channel is detectable in these data, it operates through representation and personnel rather than resource reallocation. This is consistent with Fischer (2023), who shows that board composition affects spending

Table 4: The Race-Match Channel

Outcome	Moderator	Base ATT	Interaction	p-value
Panel A: Board Composition				
BW gap	Cum. Black board share	3.56***	−33.7***	.005
HW gap	Cum. Hispanic board share	−0.17	−1.2	.384
BW gap	Cum. Hispanic board share [mismatch test]	1.41	−0.6	.834
Total rate	Cum. minority board share	0.55	−1.9*	.082
Panel B: Teacher Diversity				
BW gap	Black teacher share	3.13***	−82.0***	.0003
BW gap	Black teacher share (L1)	3.63***	−95.8***	.0001
BW gap	Hispanic teacher share [mismatch test]	1.04	+1.3	.70
Total rate	Minority teacher share (L1)	1.24**	−4.6***	.003
BW gap	St.-teacher Black gap	1.89***	−29.0**	.016

Notes: Each row reports results from a model interacting treat_post with the indicated moderator. "Base ATT" is the estimated treatment effect when the moderator is zero. "Interaction" is the coefficient on the interaction term. "L1" indicates a one-year lag of the moderator. All models include district and year fixed effects with district-clustered standard errors. $^{*}p < .10$; $^{**}p < .05$; $^{***}p < .01$.

through a separate, race-specific mechanism, and with Jackson and Mackevicius (2024), whose meta-evaluation finds that aggregate spending effects on student outcomes depend heavily on how resources are deployed.

5.5 Interpreting the Average Effect

Figure 11 synthesizes the mechanism evidence across four dimensions: the first stage (board composition dynamics around SMD adoption), the board race-match interactions, teacher mediation, and the lag structure of the board composition channel.

The mechanism evidence partially resolves the puzzle in three steps.

First, the "first stage" is race-specific: SMD elections increase Hispanic board representation by 4.0 percentage points but leave Black representation unchanged. This reflects California's demography, where Latino populations are large enough to form SMD majorities

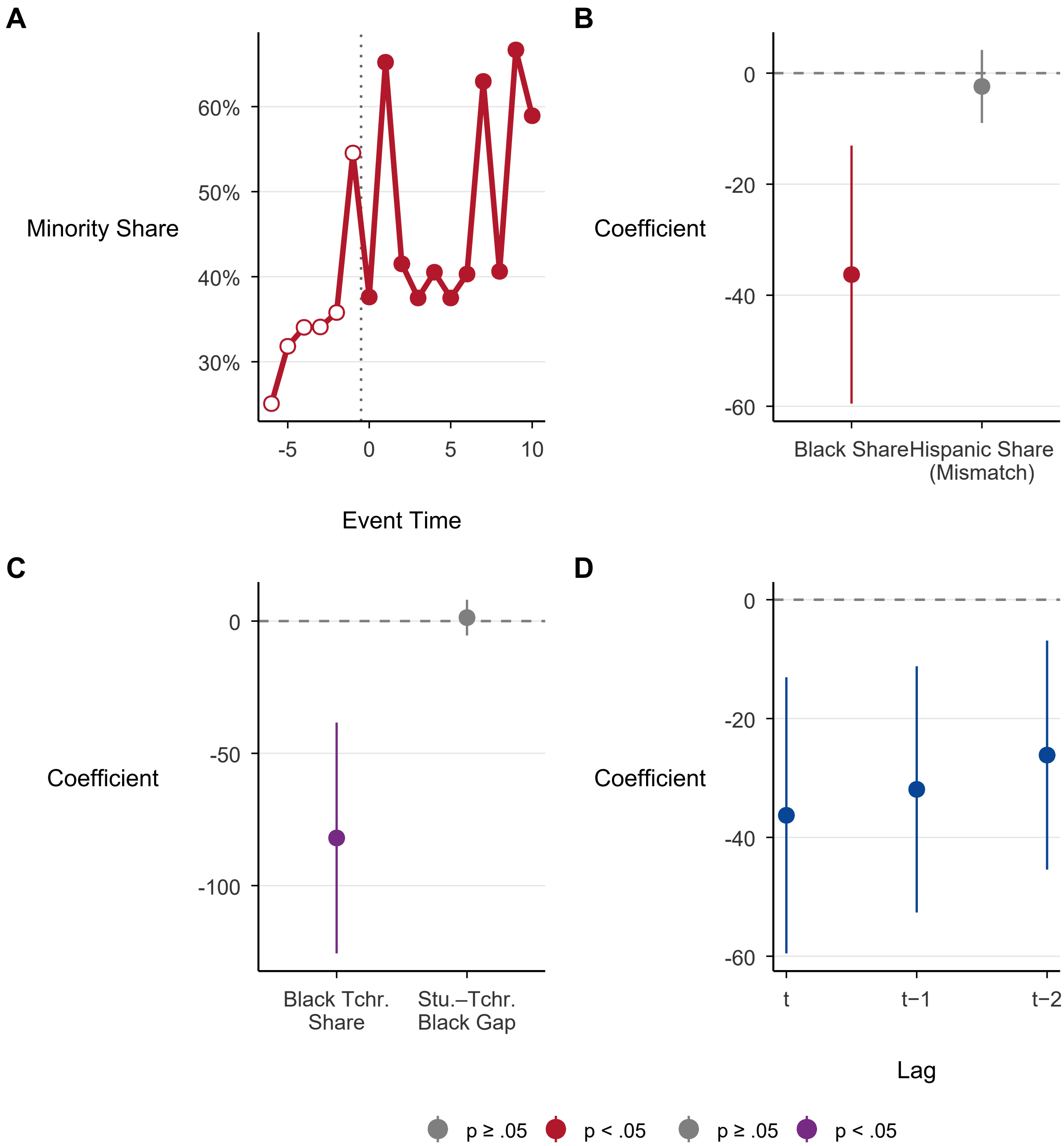


Figure 11: Mechanism Summary: Board Composition, Race-Match, and Teacher Mediation
Notes: Panel A: minority share of elected board members by event time relative to SMD adoption (hard classification). Panel B: board race-match interaction coefficients—Black and Hispanic board share interacted with SMD treatment on BW gap, with 95% CIs. Panel C: teacher mediation—Black teacher share and student-teacher Black gap interacted with SMD treatment on BW gap (note different scale from Panel B). Panel D: lag structure of the Black board share × BW gap interaction at contemporaneous, one-year, and two-year lags, with 95% CIs.

in most districts, while Black populations—constituting approximately 5–6% of the state's population—are rarely large enough to anchor a trustee area.

Second, the effects of descriptive representation are race-matched: Black board members and Black teachers are associated with reductions in the Black-White discipline gap, but Hispanic board members and Hispanic teachers are not. This finding is consistent with the theoretical framework developed in Section 2: different racial groups face different challenges in the education system, and effective advocacy requires group-specific knowledge and priorities.

Third, the average treatment effect combines these two facts. Because the vast majority of treated districts gain Hispanic but not Black representation, the average effect is dominated by districts where the BW gap is unaffected or worsens. The 17 districts that gain Black representation experience attenuated BW gap increases, but they are heavily outnumbered by the 140 districts (in the BW-gap-valid sample) that gain only Hispanic representation. The positive average BW gap effect is a composition result driven by districts that gain Hispanic but not Black representation. At the same time, the race-specific level estimates indicate higher Black suspension rates on average after adoption, so the evidence is consistent with adverse average effects concentrated in non-race-matched districts.

This interpretation carries an important nuance. The BW gap does not merely remain unchanged in districts without Black representation gains—it increases. One plausible explanation is an attention-displacement effect: as boards focus on the concerns of newly empowered Hispanic constituencies, administrative attention shifts away from Black students' needs. Flavin and Hartney (2017) document that school boards face no electoral accountability for minority outcomes under at-large elections; SMD reform does not eliminate this deficit but rather redirects it. Fischer's (2023) evidence that Hispanic board members channel resources to Hispanic-serving schools, not to schools serving other minority groups, provides the spending-domain analog. An alternative possibility is that the increasing gap reflects the reform's failure to counteract differential secular trends in Black versus White suspension rates in the absence of Black-specific advocacy. The event study's gradual ramp-up of effects is suggestive of attention-displacement, though I cannot definitively adjudicate between these explanations with the present data.

6 Discussion and Conclusion

This paper provides the first causal evidence linking electoral reform to school discipline outcomes, and in doing so suggests a boundary condition for theories of descriptive representation. The CVRA-mandated transition to single-member district elections increased minority school board representation—but primarily among Hispanic candidates, not Black candidates. As a result, the Black-White suspension gap increased by approximately 1.7 to 2.7 percentage points after reform (depending on the minimum Black enrollment threshold), while Hispanic-White gaps and total suspension rates were unaffected. The mechanism analysis is consistent with a *race-match condition*: discipline disparities are attenuated when the race of hired teachers matches the affected student group (-82 to -96 per unit Black teacher share; BY-corrected $q < .01$). A similar pattern appears for board racial composition (-34 per unit Black board share; $p = .005$), though this result is sensitive to the BISG race-classification specification and does not survive multiple-testing correction. "Minority" representation is not fungible across racial groups.

Implications for voting rights. The CVRA and similar state-level voting rights statutes are powerful tools for increasing minority representation, but their downstream effects depend on which minority groups benefit. In California, where the Latino population is large and geographically concentrated, CVRA enforcement is primarily a Latino empowerment tool. For Black communities—which are smaller and less residentially concentrated in California than in the South—the CVRA-induced transition to SMD elections may not generate the representation gains needed to reduce Black-specific disparities. The challenge is compounded by the weak accountability environment in school board elections: Payson (2017) shows that incumbents face performance accountability only in higher-turnout presidential-year elections, while in the off-cycle elections that characterize most school board races, organized interests dominate and performance signals are muted. Combined with the racial asymmetry in accountability documented by Flavin and Hartney (2017)—where only White student outcomes predict reelection—the baseline for translating representation into equitable outcomes is low. Voting rights advocates should consider complementary strategies, such as targeted candidate recruitment and support, community organizing, and direct policy advocacy, to ensure that electoral reform benefits all minority groups.

Implications for education policy. The teacher diversity results underscore the importance of race-matched teacher hiring as a lever for reducing discipline disparities (Dee 2005; Gershenson et al. 2022; Lindsay and Hart 2017). Districts that increased their Black teacher share after SMD adoption saw significant reductions in the BW discipline gap. This finding is consistent with a growing literature showing that same-race teachers reduce exclusionary discipline for Black students, likely through more culturally responsive behavioral expectations and stronger teacher-student relationships. The discipline finding parallels Fischer's (2023) result that same-race board members causally improve spending and achievement in their group's schools: both board composition and teacher composition matter, and both operate through race-specific channels. The spending channel itself is not the mediator—per-pupil spending does not moderate the BW discipline gap—suggesting that representation affects discipline through administrative signaling and personnel channels rather than resource reallocation. School boards seeking to reduce racial discipline gaps would be well served by investing in Black teacher recruitment and retention, independent of—or in addition to—changes in electoral structure.

Limitations. This study faces several important limitations that qualify the conclusions. First, the parallel trends assumption, while supported by the event study's non-monotonic pre-treatment pattern and by randomization inference, is not unambiguously validated. The joint F-test on pre-treatment coefficients rejects, and HonestDiD analysis shows that the robust confidence interval crosses zero at $M = 0.1$. While I provide multiple pieces of evidence supporting a causal interpretation—including the triple-difference estimate (+3.12; $p = .003$), randomization inference ($p = .004$), placebo tests, and the internally consistent mechanism evidence—readers should weigh these against the pre-trends concern.

Second, the BISG race predictions for board candidates, while the best available method, introduce measurement error that affects the race-match moderation results. BISG mean confidence for Black candidates is only 0.477, compared to 0.821 for Hispanic and 0.726 for White candidates. To assess sensitivity, I re-estimated the BW gap × Black board share interaction under five alternative BISG specifications: continuous posterior probabilities (baseline), hard argmax classification, posterior thresholds of 0.3 and 0.5, and a high-confidence-only filter. The interaction is significant only under the continuous-probability specification (−23.1; $p = .03$); under all four alternatives, the interaction is insignificant ($p > .30$). This sensitivity indicates that the race-match moderation result for board composition depends on the BISG specification and should be interpreted with caution. The teacher diversity results, which use administratively reported race data rather than BISG predictions, are not subject to this limitation and provide stronger evidence for the race-match channel.

Third, the study is specific to California, a state with distinctive racial demography (large Hispanic, small Black population), a particular legal framework (CVRA), and a specific political culture. The race-match condition may manifest differently in states with larger Black populations, different patterns of residential segregation, or federal Voting Rights Act coverage. Extensions to Southern states, where SMD transitions have been driven by federal law and Black populations are larger, would provide a valuable test of external validity.

Fourth, the relatively small Black population in many California school districts limits the precision and coverage of the BW gap estimates (≈60% of observations). While this coverage is sufficient for the main analysis, it means that the estimates are drawn disproportionately from the subset of California districts with larger Black communities.

Conclusion. The question "does descriptive representation reduce racial disparities?" has no single answer. It depends on whose representation increases and whose disparities are measured. In California's CVRA reform wave, descriptive representation increased for Hispanic communities but did not significantly reduce Hispanic-specific disparities. But it did not increase for Black communities, and Black-White disparities widened. The lesson is not that descriptive representation fails, but that it succeeds only when matched to the specific group it represents. Electoral reform is a necessary but insufficient condition for equitable governance; the specific racial composition of the resulting representation determines which communities benefit.

References

Abott, Carolyn, and Asya Magazinnik. 2020. "At-Large Elections and Minority Representation in Local Government." *American Journal of Political Science* 64(3): 717–733.

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. 2021. "Synthetic Difference-in-Differences." *American Economic Review* 111(12): 4088–4118.

Benjamini, Yoav, and Daniel Yekutieli. 2001. "The Control of the False Discovery Rate in Multiple Testing Under Dependency." *Annals of Statistics* 29(4): 1165–1188.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2024. "Revisiting Event-Study Designs: Robust and Efficient Estimation." *Review of Economic Studies* 91(6): 3253–3285.

California Department of Education. 2024. Suspension and Expulsion Data. Data file California Department of Education.

California Elections Data Archive. 2024. School Board Election Results. Data file California State University, Sacramento.

Callaway, Brantly, and Pedro H.C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics* 225(2): 200–230.

Collingwood, Loren, and Sean Long. 2021. "Can States Promote Minority Representation? Assessing the Effects of the California Voting Rights Act." *Urban Affairs Review* 57(3): 731–762.

Davidson, Chandler, and Bernard Grofman. 1994. *Quiet Revolution in the South: The Impact of the Voting Rights Act, 1965–1990.* Princeton: Princeton University Press.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110(9): 2964–2996.

Dee, Thomas S. 2005. "A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?" *American Economic Review* 95(2): 158–165.

Engstrom, Richard L., and Michael D. McDonald. 1981. "The Election of Blacks to City Councils: Clarifying the Impact of Electoral Arrangements on the Seats/Population Relationship." *American Political Science Review* 75(2): 344–354.

Fischer, Brett. 2023. "No Spending without Representation: School Boards and the Racial Gap in Education Finance." *American Economic Journal: Economic Policy* 15(2): 198–235.

Flavin, Patrick, and Michael T. Hartney. 2017. "Racial Inequality in Democratic Accountability: Evidence from Retrospective Voting in Local Elections." *American Journal of Political Science* 61(3): 684–697.

Gershenson, Seth, Cassandra M.D. Hart, Joshua Hyman, Constance Lindsay, and Nicholas W. Papageorge. 2022. "The Long-Run Impacts of Same-Race Teachers." *American Economic Journal: Economic Policy* 14(4): 300–342.

Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics* 225(2): 254–277.

Imai, Kosuke, and Kabir Khanna. 2016. "Improving Ecological Inference by Predicting Individual Ethnicity from Voter Registration Records." *Political Analysis* 24(2): 263–272.

Jackson, C. Kirabo, and Claire Mackevicius. 2024. "What Impacts Can We Expect from School Spending Policy? Evidence from Evaluations in the United States." *American Economic Journal: Applied Economics* 16(1): 412–446.

Kogan, Vladimir, Stéphane Lavertu, and Zachary Peskowitz. 2021. "How Does Minority Political Representation Affect School District Administration and Student Outcomes?" *American Journal of Political Science* 65(3): 699–716.

Lal, Apoorva, and Daniel M. Thompson. 2024. "Did Private Election Administration Funding Advantage Democrats in 2020?" *Proceedings of the National Academy of Sciences* 121(22): e2317563121.

Lawyers' Committee for Civil Rights of the San Francisco Bay Area. 2014. Voting Rights Database: California School Districts Transitioning to By-Trustee Area Elections. Database LCCR.

Leal, David L., Valerie Martinez-Ebers, and Kenneth J. Meier. 2004. "The Politics of Latino Education: The Biases of At-Large Elections." *Journal of Politics* 66(4): 1224–1244.

Lindsay, Constance A., and Cassandra M.D. Hart. 2017. "Exposure to Same-Race Teachers and Student Disciplinary Outcomes for Black Students in North Carolina." *Educational Evaluation and Policy Analysis* 39(3): 485–510.

Mansbridge, Jane. 1999. "Should Blacks Represent Blacks and Women Represent Women? A Contingent "Yes"." *Journal of Politics* 61(3): 628–657.

Marschall, Melissa J., Anirudh V.S. Ruhil, and Paru R. Shah. 2010. "The New Racial Calculus: Electoral Institutions and Black Representation in Local Legislatures." *American Journal of Political Science* 54(1): 107–124.

Meier, Kenneth J., and Amanda Rutherford. 2016. *The Politics of African-American Education: Representation, Partisanship, and Educational Equity.* New York: Cambridge University Press.

Meier, Kenneth J., and Robert E. England. 1984. "Black Representation and Educational Policy: Are They Related?" *American Political Science Review* 78(2): 392–403.

Meier, Kenneth J., Joseph Stewart, and Robert E. England. 1989. *Race, Class, and Education: The Politics of Second-Generation Discrimination.* Madison: University of Wisconsin Press.

Morris, Edward W., and Brea L. Perry. 2016. "The Punishment Gap: School Suspension and Racial Disparities in Achievement." *Social Problems* 63(1): 68–86.

Payson, Julia A. 2017. "When Are Local Incumbents Held Accountable for Government Performance? Evidence from U.S. School Districts." *Legislative Studies Quarterly* 42(3): 421–448.

Pitkin, Hanna Fenichel. 1967. *The Concept of Representation.* Berkeley: University of California Press.

Rambachan, Ashesh, and Jonathan Roth. 2023. "A More Credible Approach to Parallel Trends." *Review of Economic Studies* 90(5): 2555–2591.

Reardon, Sean F., Andrew D. Ho, Benjamin R. Shear, Erin M. Fahle, Demetra Kalogrides, and J. Saliba. 2026. Stanford Education Data Archive (Version 6.0). Data release Stanford University.

Skiba, Russell J., Robert H. Horner, Choong-Geun Chung, M. Karega Rausch, Seth L. May, and Tary Tobin. 2011. "Race Is Not Neutral: A National Investigation of African American and Latino Disproportionality in School Discipline." *School Psychology Review* 40(1): 85–107.

Skiba, Russell J., Robert S. Michael, Abra Carroll Nardo, and Reece L. Peterson. 2002. "The Color of Discipline: Sources of Racial and Gender Disproportionality in School Punishment." *Urban Review* 34(4): 317–342.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics* 225(2): 175–199.

Trebbi, Francesco, Philippe Aghion, and Alberto Alesina. 2008. "Electoral Rules and Minority Representation in U.S. Cities." *Quarterly Journal of Economics* 123(1): 325–357.

Trounstine, Jessica, and Melody E. Valdini. 2008. "The Context Matters: The Effects of Single-Member versus At-Large Districts on City Council Diversity." *American Journal of Political Science* 52(3): 554–569.

Xu, Yiqing. 2017. "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models." *Political Analysis* 25(1): 57–76.

Online Appendix

A Data Construction Details

This appendix documents the construction of the analysis dataset, with detailed coverage statistics and data quality diagnostics for each data source.

A.1 CEDA Election Data

The California Elections Data Archive (CEDA) provides candidate-level records for school board elections from 1996 through 2023. I load annual Excel files, harmonize variable names across years, and standardize district names for cross-referencing with CDE data. The `area` field, which records trustee area identifiers for district-based elections, is the basis for identifying SMD systems.

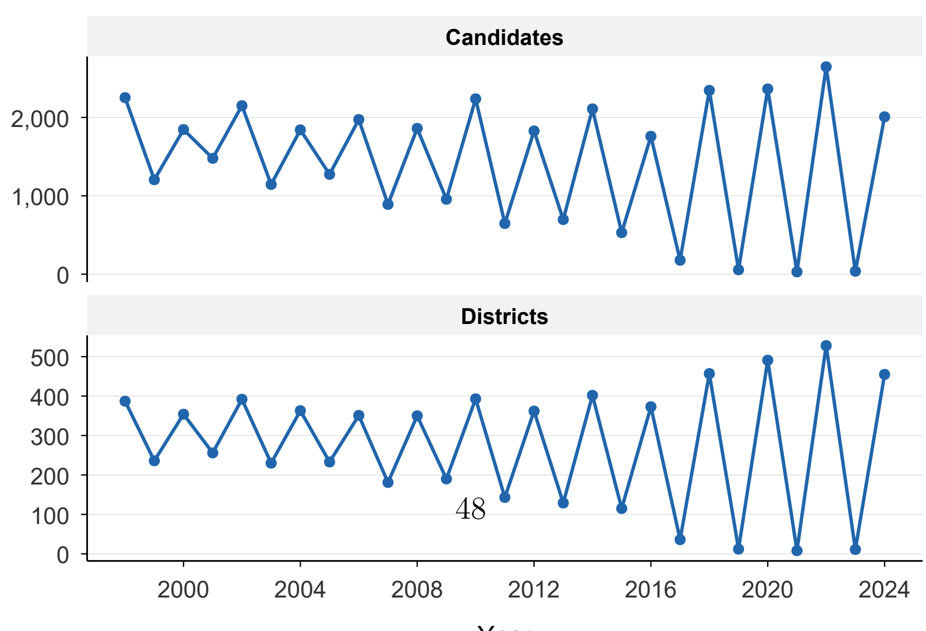
Table A1: Overview of All Data Sources Used in Analysis

Data Source	Years	N Observations	N Districts	Key Variables
CEDA Election Data	1998–2024	38,340	1,437	Candidates, electoral system, winners
CDE Suspension Data	2011–2024	542,998	1,089	Suspension rates by race, gaps
CDE Expulsion Data	2011–2024	542,998	1,877	Expulsion rates by race, gaps
Analysis Dataset	2011–2024	10,786	788	Merged outcomes + treatment

Note: Suspension and expulsion data have different district coverage because CDE reporting requirements differ across discipline categories.

Table A2: CEDA Election Data Coverage by Year

Year	Districts	Candidates	Avg. Cand./Dist.	Ward Elections	% Ward
1998	387	2,252	5.8	504	22.4
1999	236	1,205	5.1	158	13.1
2000	354	1,845	5.2	417	22.6
2001	256	1,478	5.8	247	16.7
2002	392	2,150	5.5	399	18.6
2003	230	1,145	5.0	140	12.2
2004	363	1,841	5.1	388	21.1
2005	233	1,274	5.5	196	15.4
2006	351	1,973	5.6	379	19.2
2007	181	891	4.9	199	22.3
2008	350	1,860	5.3	375	20.2
2009	190	956	5.0	141	14.7
2010	393	2,237	5.7	471	21.1
2011	143	647	4.5	84	13.0
2012	362	1,827	5.0	519	28.4
2013	129	697	5.4	119	17.1
2014	402	2,109	5.2	563	26.7
2015	115	530	4.6	152	28.7
2016	373	1,759	4.7	571	32.5
2017	36	178	4.9	66	37.1
2018	457	2,345	5.1	864	36.8
2019	12	56	4.7	33	58.9
2020	491	2,363	4.8	1,133	47.9
2021	8	30	3.8	10	33.3
2022	528	2,646	5.0	1,364	51.5
2023	11	38	3.5	31	81.6
2024	455	2,008	4.4	1,195	59.5



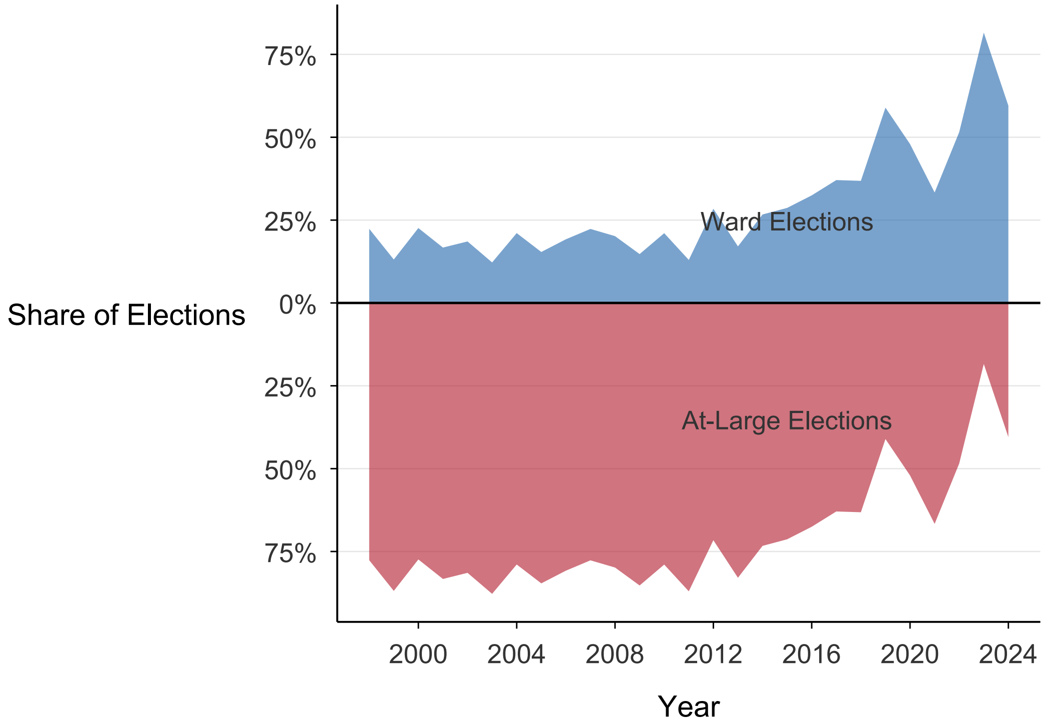


Figure A2: Electoral System Adoption Trends

Notes: Ward (SMD) elections shown above axis; at-large elections below axis. Each series shows the share of all CEDA-recorded school board elections in that year using the given electoral system.

A.2 District Characteristics

The table below summarizes the CEDA district universe and electoral system classifications. Of the districts observed in CEDA, a substantial majority used at-large elections throughout the sample period, while a smaller share transitioned to single-member districts. Figure A3 shows the distribution of years covered per district; the panel is nearly balanced, with most districts observed in all 14 years.

Table A3: CEDA District Summary Statistics

Category	Value
Total CEDA districts	1,437
Classified districts	1,568
Years of coverage (median)	4.0
Years of coverage (mean)	5.2
Districts with 10+ years	301
Districts with 20+ years	0
Ever used ward elections	590
Always ward elections	238
Always at-large elections	978
Switched electoral systems	352

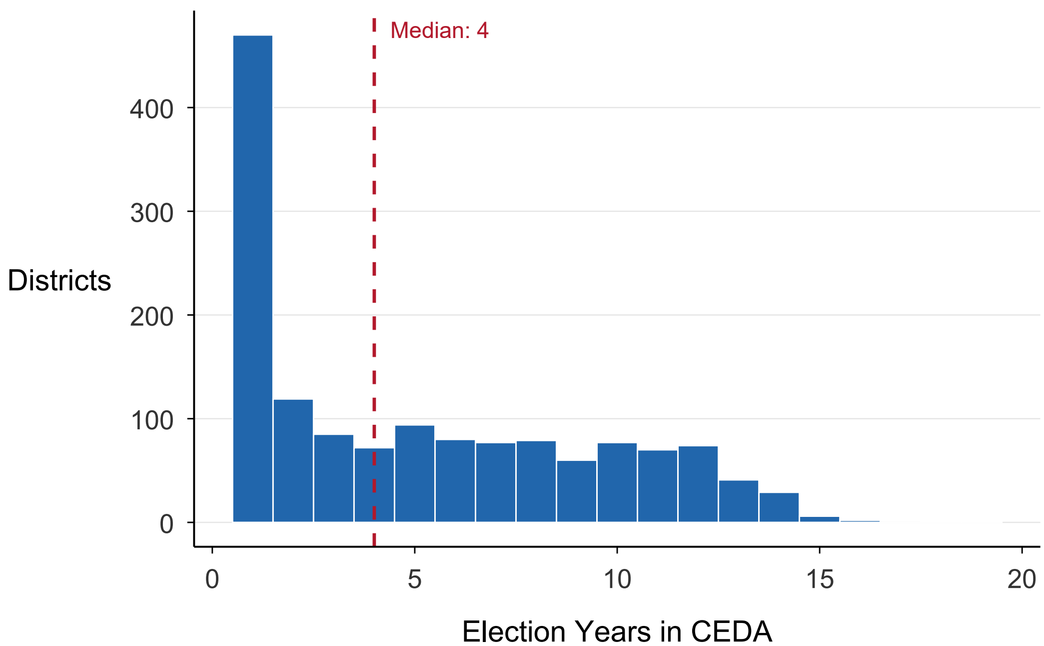


Figure A3: Distribution of Years Covered per District

A.3 Candidate Race/Ethnicity Predictions

I predict candidates' race/ethnicity using Bayesian Improved Surname Geocoding (BISG) via the `wru` package (Imai and Khanna 2016).

Table A4: BISG Race/Ethnicity Predictions for School Board Candidates

Race/Ethnicity	Candidates	%	Mean Confidence
White	25,334	66.1	0.726
Hispanic	10,708	27.9	0.821
Asian	1,676	4.4	0.848
Black	569	1.5	0.477
Other	52	0.1	0.578
Total	38,340	100.0	0.754

Notes: Race/ethnicity predicted via Bayesian Improved Surname Geocoding (BISG) using the `wru` package (Imai and Khanna 2016). Confidence is the maximum posterior probability across racial categories (argmax classification). Mean predicted minority probability: 48.6%.

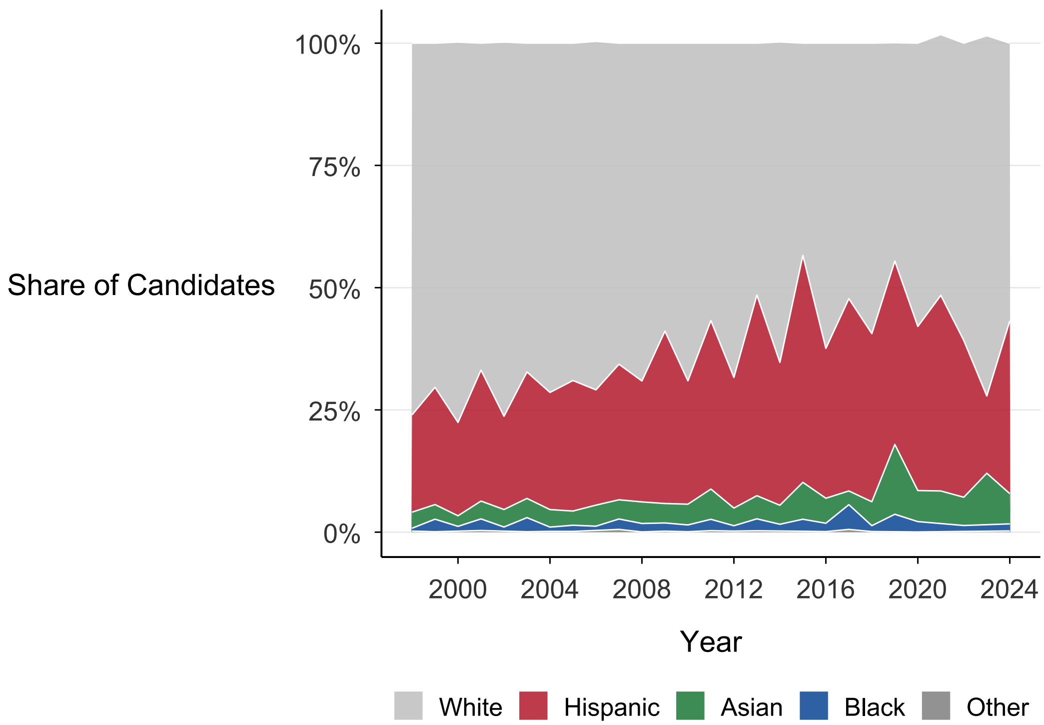


Figure A4: Racial/Ethnic Composition of Candidates Over Time

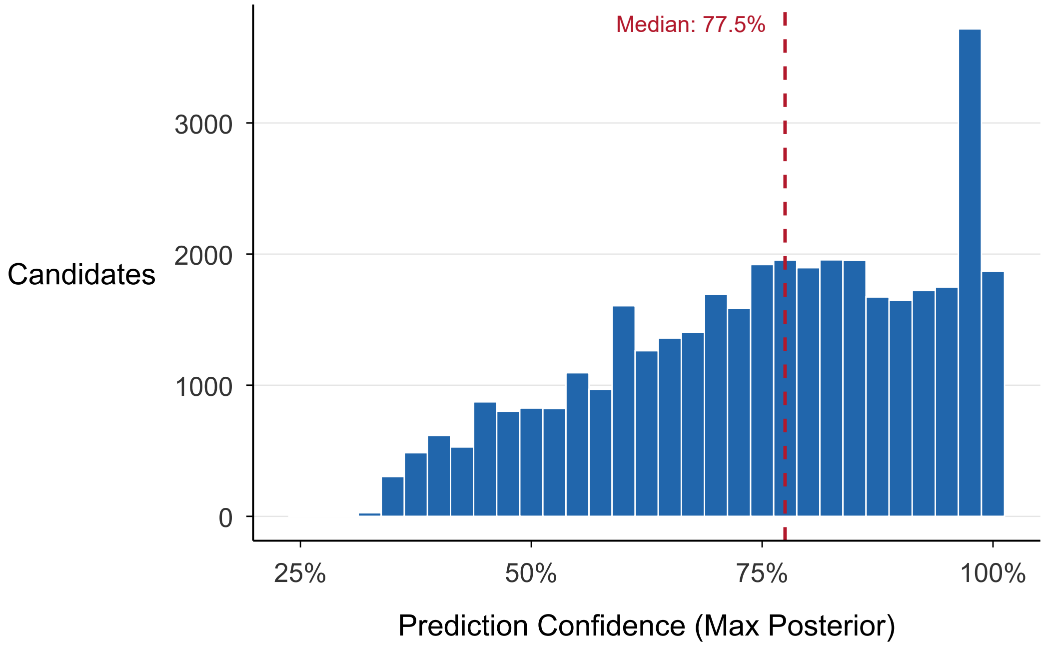


Figure A5: Distribution of BISG Prediction Confidence

A.4 CDE Suspension Data

Suspension data are downloaded from the California Department of Education for academic years 2010–11 through 2023–24. The coverage table below reports the number of districts with suspension data in each year; coverage is stable at approximately 1,000 districts per year. Figure A6 plots the mean BW and HW suspension rate gaps over time, separately for treated and control districts, while Figure A7 shows the cross-sectional distribution of gaps.

Table A5: Suspension Data Coverage by Year

		Valid Gap		Mean Rate (p.p.)		
Year	N	BW	HW	BW Gap	HW Gap	Black
2011	783	461	693	12.0	0.4	21.8
2012	778	460	707	10.6	0.1	18.7
2013	773	468	717	9.1	0.0	16.1
2014	772	467	713	8.4	-0.2	14.5
2015	772	458	717	7.8	-0.4	13.9
2016	770	470	719	7.7	-0.5	13.9
2017	770	464	720	7.7	0.0	12.8
2018	767	455	718	6.8	-0.2	11.9
— *COVID years 2019–20, 2020–21 excluded* —						
2021	768	454	719	6.3	0.2	10.8
2022	766	447	718	7.2	0.0	12.2
2023	763	444	716	6.9	-0.2	11.7
2024	765	442	716	6.2	-0.3	10.6

Notes: Non-COVID district-year observations from the analysis panel. N = districts observed. Valid Gap = districts with sufficient enrollment to compute gap. Gaps and rates in percentage points.

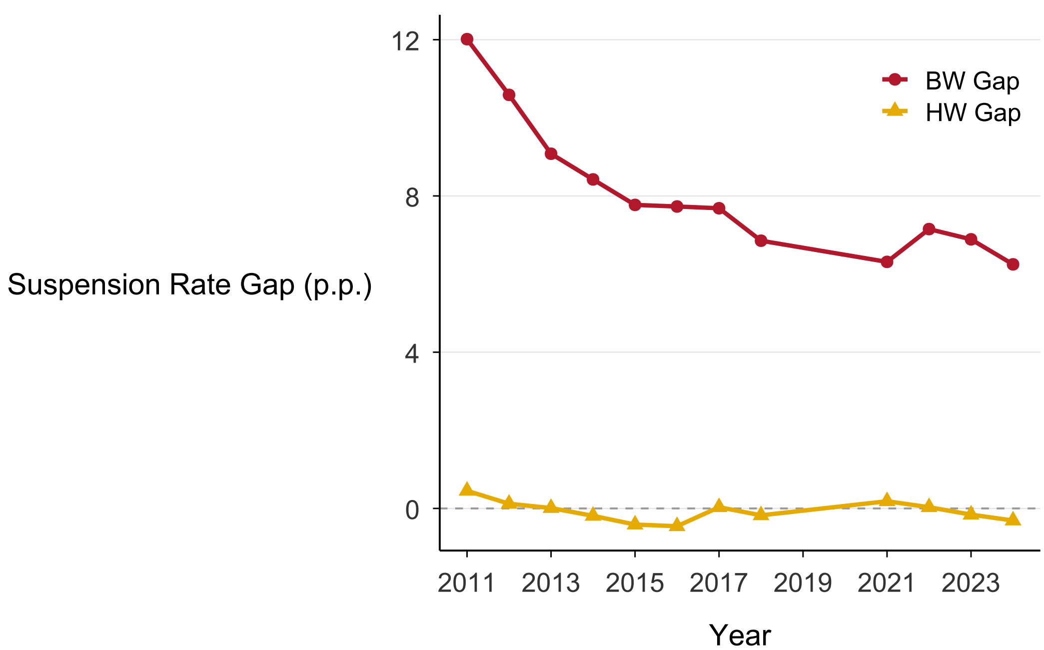


Figure A6: Mean Suspension Rate Gaps Over Time

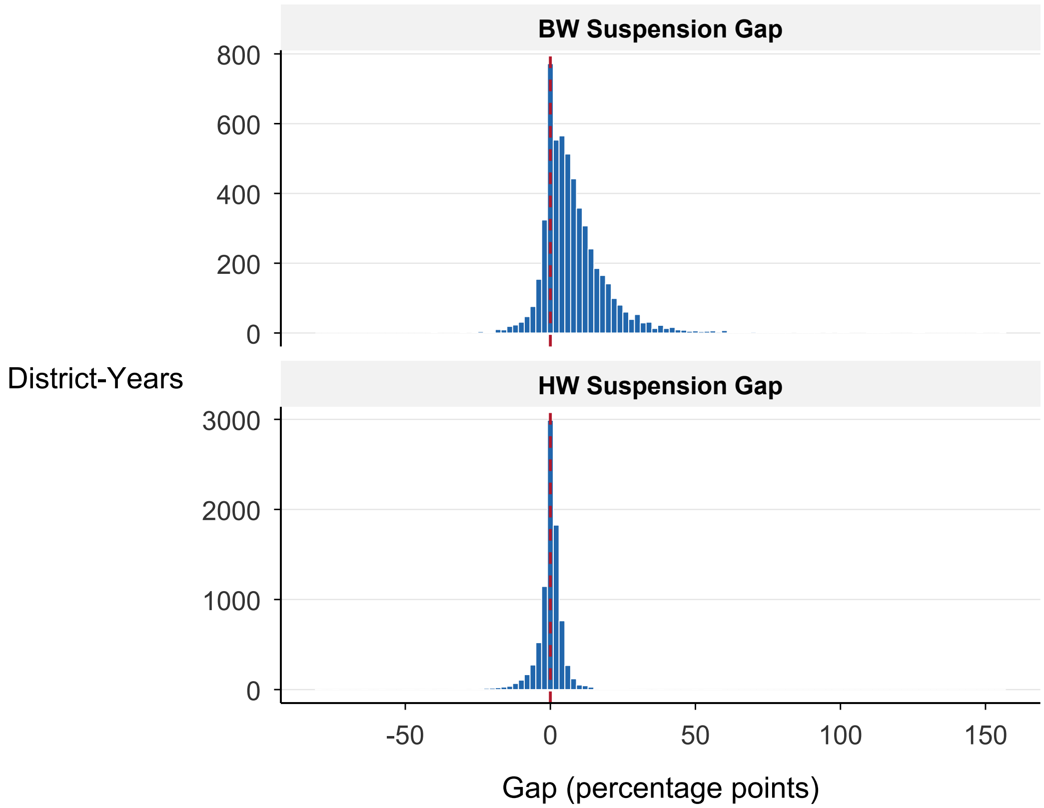


Figure A7: Distribution of Racial Suspension Gaps

Table A6: Summary Statistics: Suspension Rates by Race/Ethnicity

Race	N	Mean	SD	Median	25th	75th
Black	5,500	14.1	15.9	10.0	4.0	19.2
Hispanic	8,757	5.8	6.4	4.4	1.9	7.6
White	8,871	5.9	7.4	3.8	1.4	7.7
Asian	5,836	2.3	4.4	1.0	0.0	2.6
Total	9,155	5.9	6.8	4.1	1.7	7.8

Notes: Suspension rates in percentage points (suspensions per 100 enrolled students). N = non-COVID district-year observations with non-missing enrollment for the given group. Total includes all students regardless of race.

A.5 Expulsion Data

Expulsion data come from the same CDE source and cover the same 2011–2024 window. Coverage is stable at approximately 1,020 districts per year throughout the sample period.

Expulsion rates are an order of magnitude smaller than suspension rates and are used as a secondary outcome. Figure A8 shows the relationship between suspension and expulsion BW gaps at the district-year level.

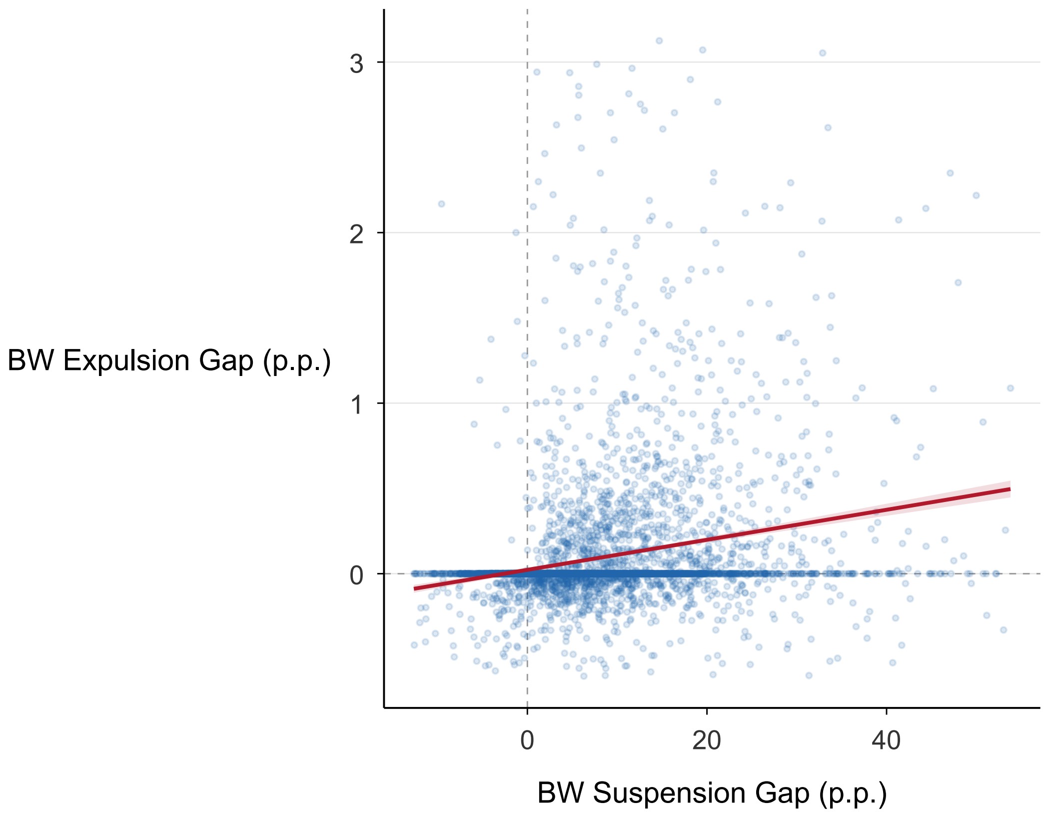


Figure A8: Relationship Between Suspension and Expulsion Gaps

Table A7: Expulsion Data Coverage by Year

Year	Districts with Data
2011	1,044
2012	1,038
2013	1,028
2014	1,022
2015	1,024
2016	1,023
2017	1,027
2018	1,039
2019	1,033
2020	1,029
2021	1,020
2022	1,019
2023	1,010
2024	1,015

A.6 Treatment Timing

The table below reports the distribution of treatment cohorts—the year in which each district first held a single-member district election. The largest cohorts are 2022 (51 districts), 2020 (44), and 2018 (34), reflecting the acceleration of CVRA-induced transitions in the late 2010s. Figure A9 visualizes the treatment stagger across districts and years, and Figure A10 shows the distribution of post-treatment exposure (number of post-treatment years observed).

Table A8: Distribution of Treatment Timing (Ward Election Adoption)

Cohort Year	Districts	%	Cum. %
2004	1	0.4	0.4
2005	3	1.3	1.8
2006	2	0.9	2.7
2007	1	0.4	3.1
2008	3	1.3	4.5
2009	3	1.3	5.8
2010	3	1.3	7.2
2011	2	0.9	8.1
2012	19	8.5	16.6
2013	5	2.2	18.8
2014	19	8.5	27.4
2015	8	3.6	30.9
2016	22	9.9	40.8
2017	2	0.9	41.7
2018	34	15.2	57.0
2020	44	19.7	76.7
2022	51	22.9	99.6
2023	1	0.4	100.0
Total	223	100.0	

Notes: Cohort year = year of first single-member district election. Cohorts before 2011 adopted SMD prior to the analysis window; 16 of 223 treated districts fall in this group. The analysis window covers 2011–2024 (excluding COVID years 2019–20, 2020–21).

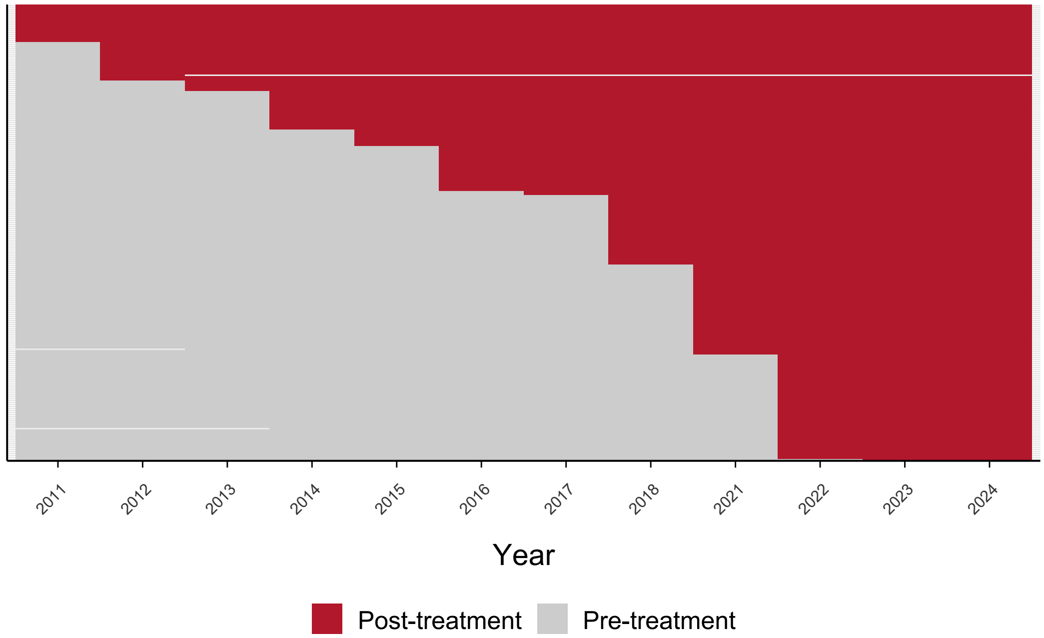


Figure A9: Treatment Stagger Visualization

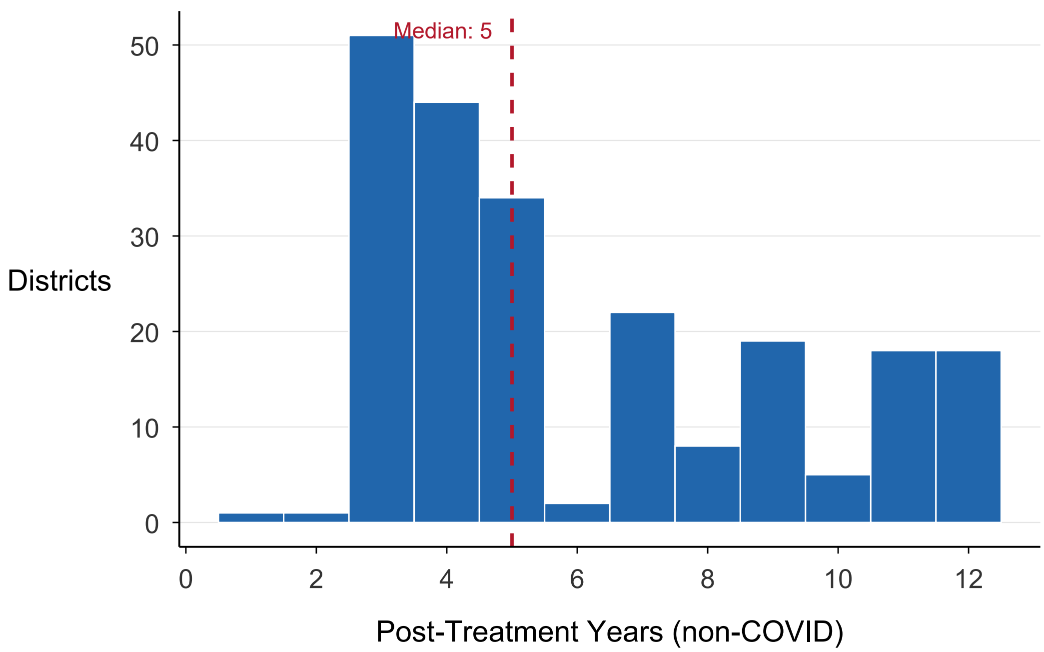


Figure A10: Distribution of Post-Treatment Exposure

A.7 Analysis Sample

The tables below summarize the composition of the final analysis panel after merging treatment and outcome data and excluding COVID years. The panel covers 788 districts over 2011–2024 (9,439 non-COVID district-year observations). The balance table compares pre-treatment means for treated and control districts; treated districts are larger, more diverse, and start with higher discipline gaps. Figure A11 maps sample coverage across California counties. The final table documents missing data patterns; the BW gap is available for approximately 60% of observations (due to small Black enrollment in many California districts), while the HW gap has over 90% coverage.

Table A9: Final Analysis Sample Composition

Category	Value
Total observations	9,247
Unique districts	788
Years covered	2011 - 2024
Treated district-years ($SMD_{dt} = 1$)	1,367
Untreated district-years ($SMD_{dt} = 0$)	7,880
Pre-treatment observations	7,880
Post-treatment observations	1,367
Districts with valid BW gap	528
Districts with valid HW gap	755
Avg years per district	11.7

Table A10: Balance Table: Pre-Treatment Characteristics

	Control		Treated		
Variable	Mean	SD	Mean	SD	Norm. Diff
Suspension rate (total)	5.60	7.25	6.96	6.39	0.20
Suspension rate (Black)	13.31	17.31	15.98	15.18	0.16
Suspension rate (Hispanic)	5.46	6.88	7.16	5.73	0.27
Suspension rate (White)	5.83	7.96	6.14	5.85	0.04
BW suspension gap	7.12	13.18	9.78	11.77	0.21
HW suspension gap	-0.37	5.58	1.03	3.86	0.29
Expulsion rate (total)	0.06	0.22	0.10	0.17	0.21
BW expulsion gap	0.12	0.76	0.15	0.55	0.05
HW expulsion gap	0.01	0.32	0.04	0.15	0.11
Enrollment (total)	3,434	7,070	11,296	10,803	0.86

Notes: Pre-treatment means for control districts (all years) and treated districts (pre-treatment years only). Rates and gaps in percentage points. Normalized difference = $(\bar{x}_T - \bar{x}_C)/\sqrt{(s_T^2 + s_C^2)/2}$. Values > $|0.25|$ suggest meaningful imbalance.

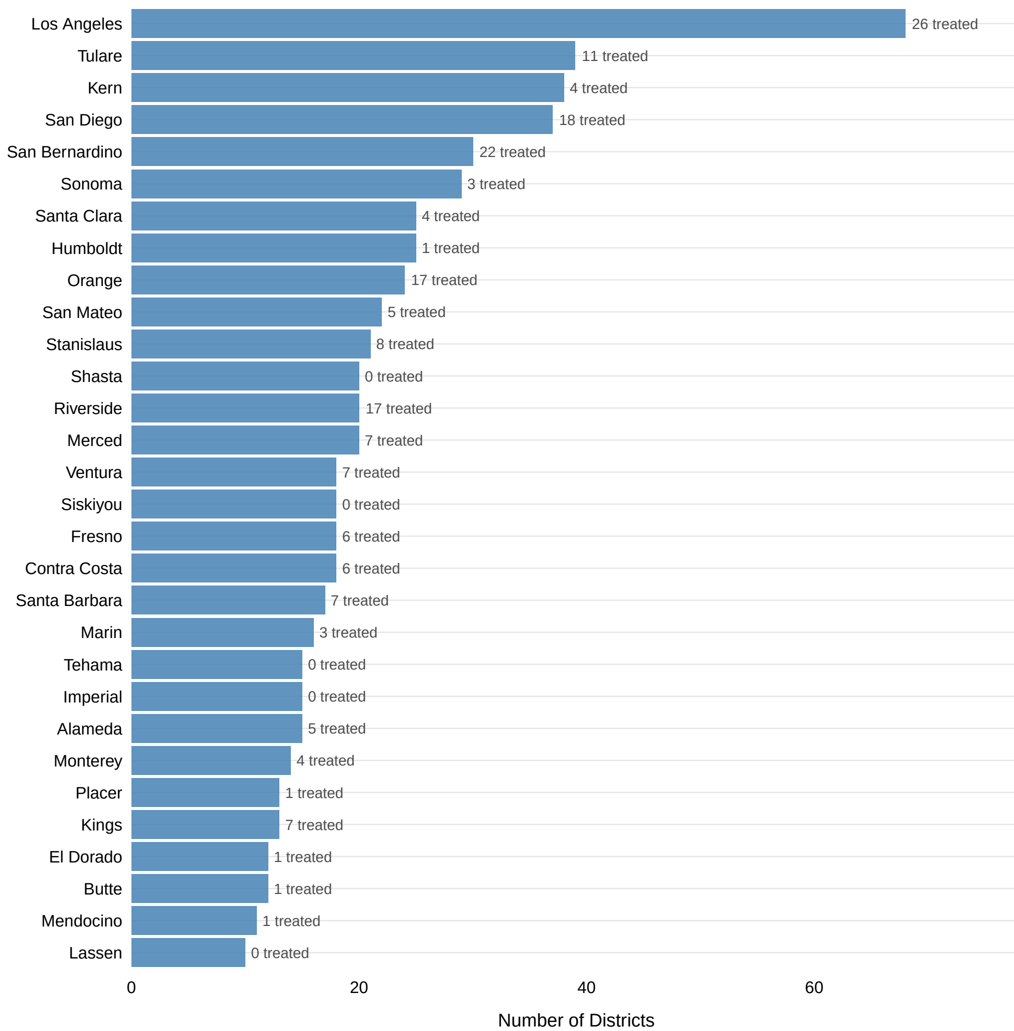


Figure A11: Sample Coverage by County

Table A11: Missing Data Patterns in Analysis Dataset

Variable	N Valid	N Missing	% Missing
Suspension rate (total)	9,155	92	1.0
Suspension rate (Black)	5,500	3,747	40.5
Suspension rate (Hispanic)	8,757	490	5.3
Suspension rate (White)	8,871	376	4.1
Suspension rate (Asian)	5,836	3,411	36.9
BW suspension gap	5,490	3,757	40.6
HW suspension gap	8,573	674	7.3
Expulsion rate (total)	8,580	667	7.2
BW expulsion gap	5,182	4,065	44.0
HW expulsion gap	8,039	1,208	13.1
Enrollment (total)	9,247	0	0.0

Notes: Out of 9,247 non-COVID district-year observations. BW gap missingness is driven by small Black enrollment (districts with <5 Black students are coded as missing). HW gap has higher coverage due to California's large Hispanic population.

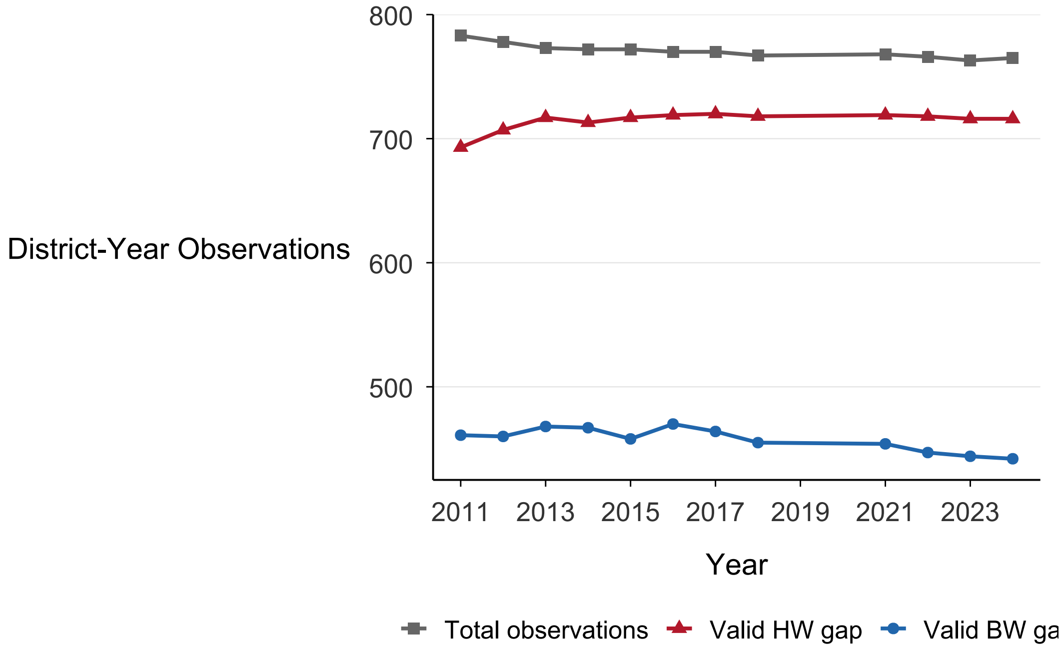


Figure A12: Data Availability Over Time

B Additional Descriptive Figures

This appendix presents supplementary descriptive evidence on suspension rate trends. Figure A13 plots race-specific suspension rates (Black, Hispanic, White) separately for treated and control districts, showing that the treated–control divergence is concentrated among Black students. Figure A14 presents raw HW gap trends, which show no clear divergence—consistent with the null DiD estimates. Figures A15 and A16 restrict the sample to pre-treatment years for each cohort, providing a direct visual check of the parallel trends assumption.

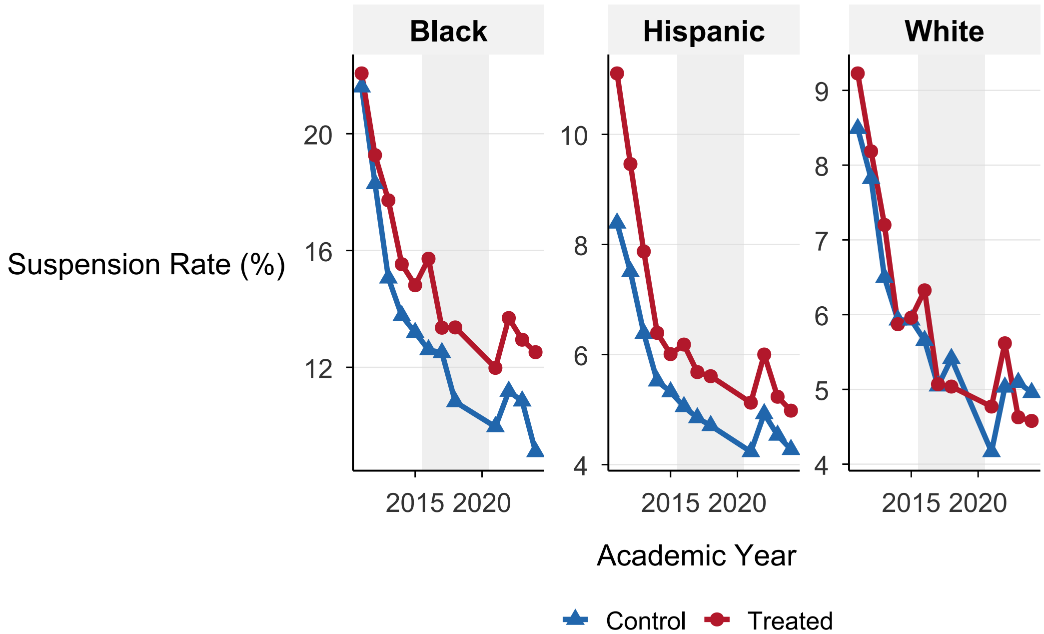


Figure A13: Race-Specific Suspension Rates by Treatment Group

Notes: Mean suspension rates by race (Black, Hispanic, White) for treated (solid) and control (dashed) districts. The shaded band marks the modal treatment window. Black rates are highest and show the largest treated-control divergence. COVID years excluded.

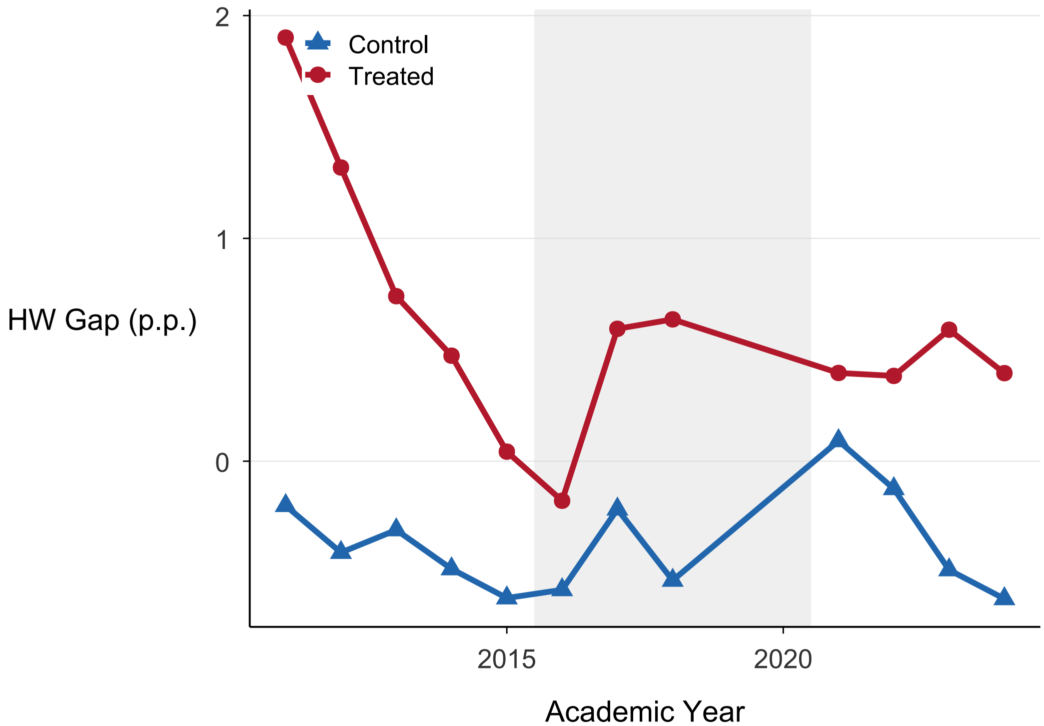


Figure A14: Raw Trends: Hispanic-White Suspension Rate Gap

Notes: Annual mean Hispanic-White suspension rate gap (percentage points) for treated and control districts. Unlike the BW gap (Figure 2), the HW gap shows no clear divergence after the modal treatment window, consistent with the null DiD estimates.

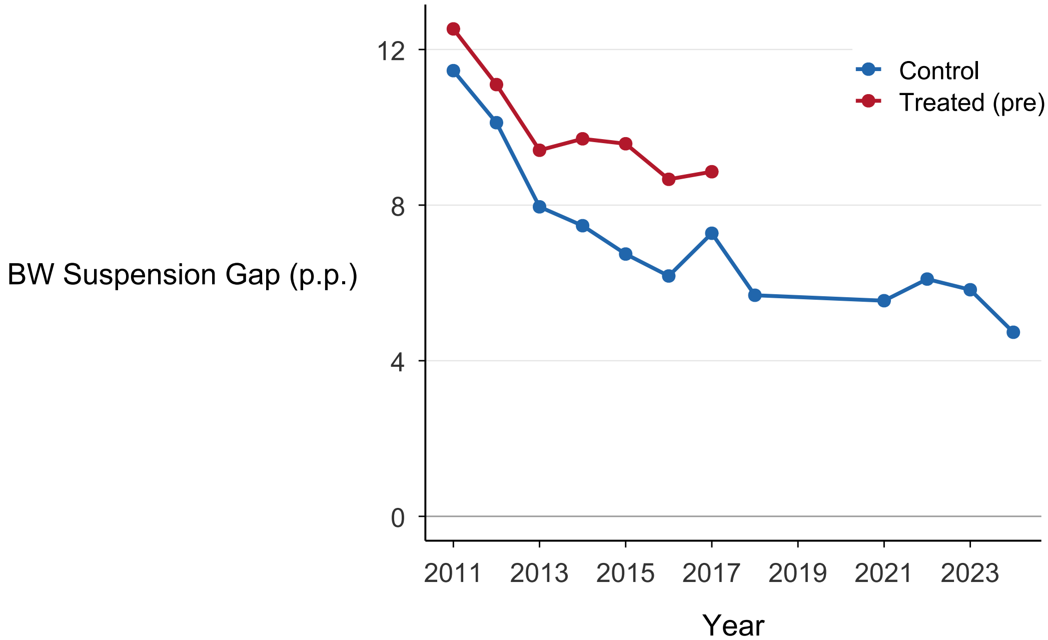


Figure A15: Pre-Treatment Trends: BW Suspension Gap

Notes: Pre-treatment BW suspension gap trends for treated and control districts, restricted to years before each cohort's first treatment year. Both groups show declining gaps in the early period, with no systematic divergence.

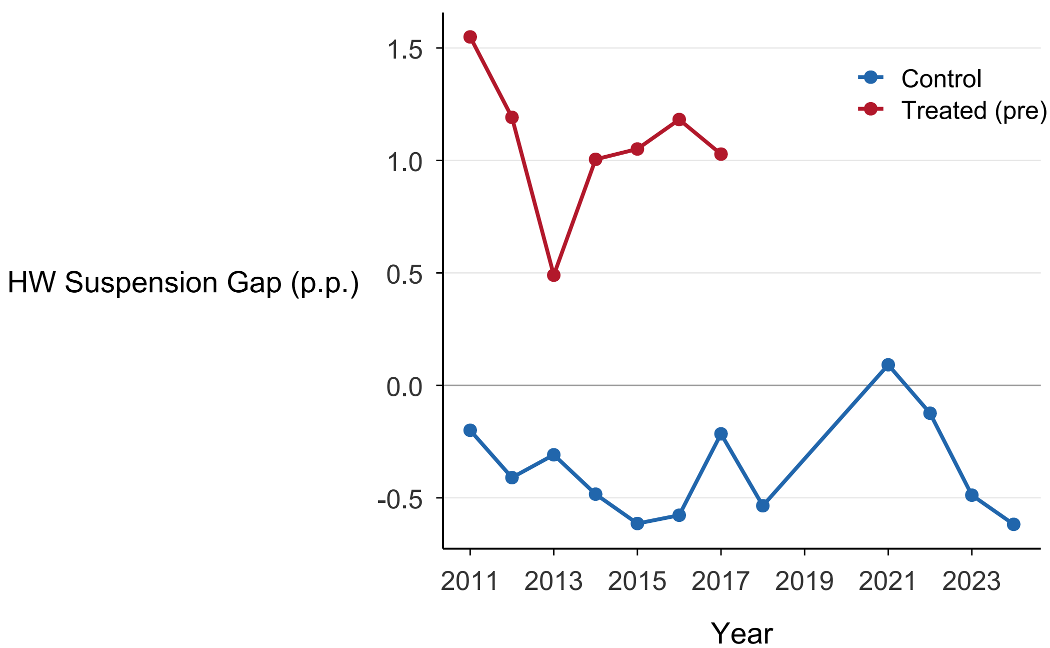


Figure A16: Pre-Treatment Trends: HW Suspension Gap

Notes: Pre-treatment HW suspension gap trends for treated and control districts.

C Additional Event Studies

This appendix reports event study estimates for outcomes not shown in the main text. Figure A17 presents the HW gap event study across all three heterogeneity-robust estimators; all coefficients are near zero both before and after treatment. Figure A18 shows the same for the total suspension rate. Figures A19 and A20 provide diagnostic plots for the synthetic DiD estimator, including the period-by-period gap between observed and synthetic control units and cohort-specific SDID estimates.

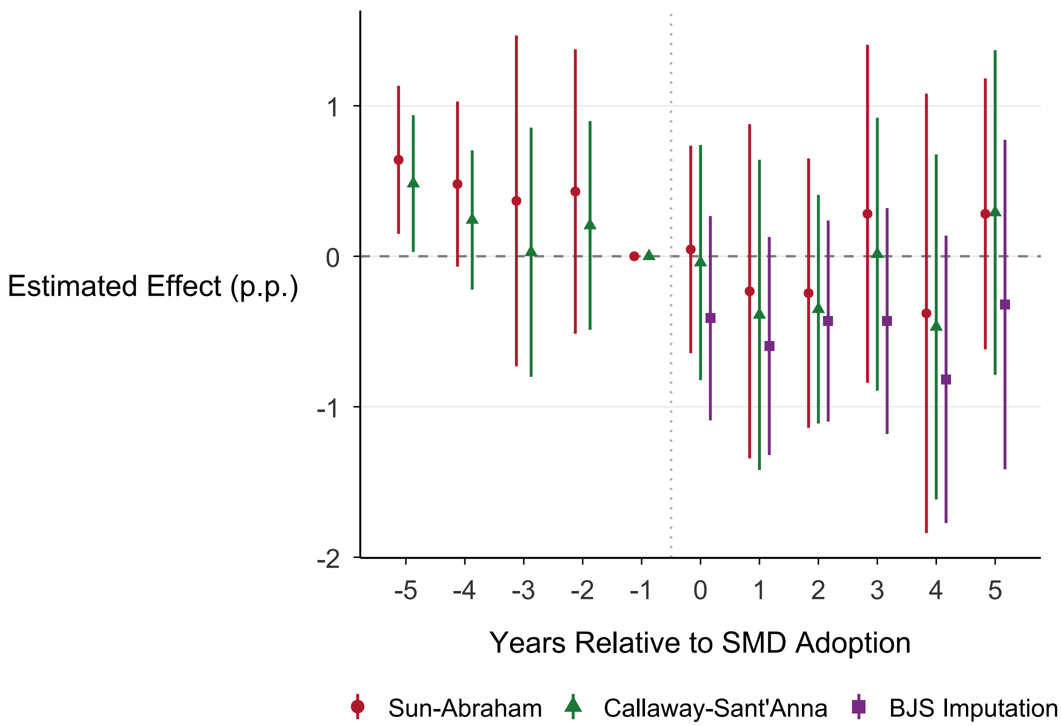


Figure A17: Event Study: Hispanic-White Suspension Gap

Notes: Sun-Abraham, Callaway-Sant'Anna, and BJS event study estimates for the Hispanic-White suspension rate gap. The dependent variable is in percentage points. Whiskers show 95% confidence intervals. COVID years excluded.

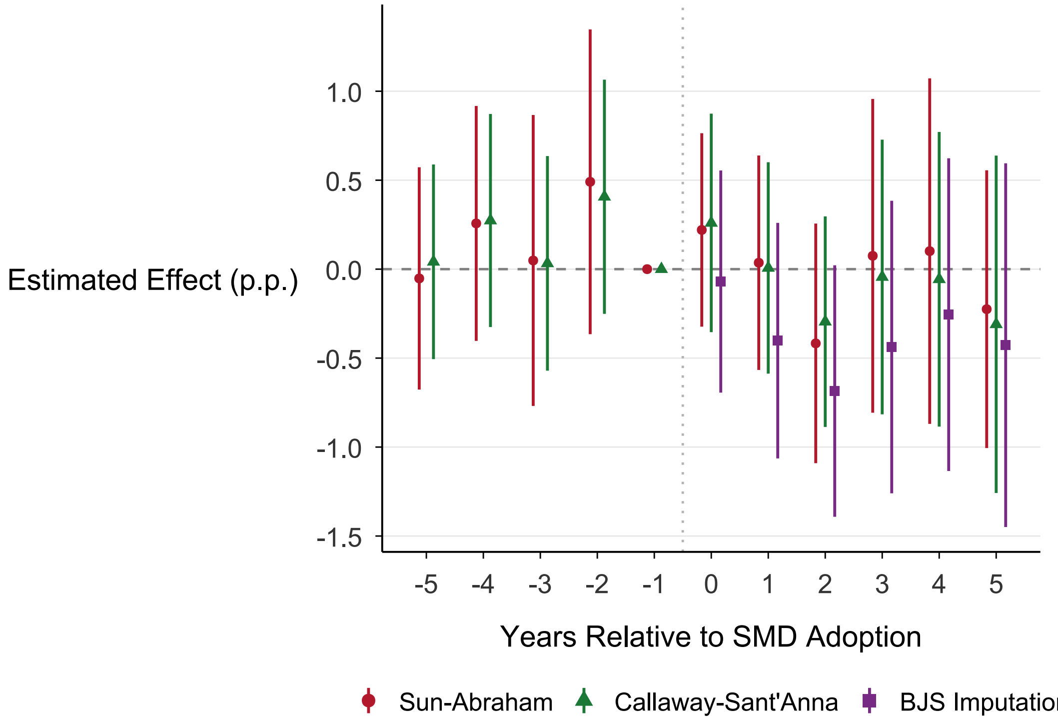


Figure A18: Event Study: Total Suspension Rate

Notes: Sun-Abraham, Callaway-Sant'Anna, and BJS event study estimates for the total suspension rate per 100 students. Whiskers show 95% confidence intervals. COVID years excluded.

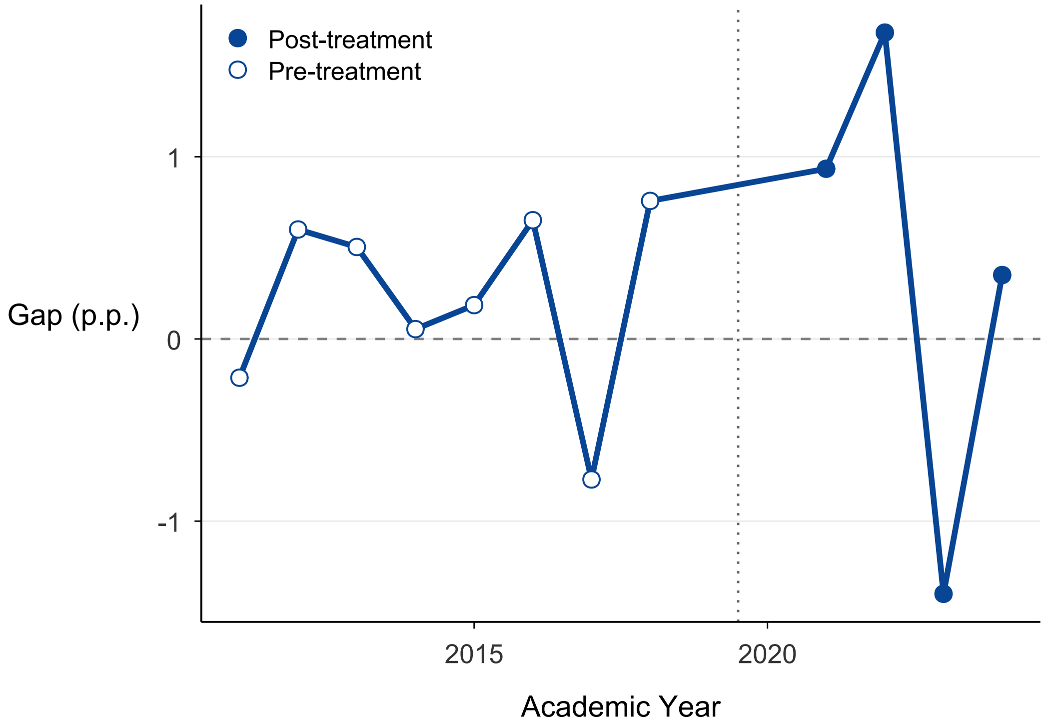


Figure A19: SDID Gap Plot: BW Suspension Gap (Diagnostic)
Notes: Period-by-period difference between observed and SDID synthetic control for the BW gap (largest treatment cohort, 2020). Pre-treatment residuals exhibit meaningful fluctuation (±0.8 p.p.), reflecting the difficulty of constructing a close synthetic match for the BW gap—consistent with the smaller, statistically insignificant SDID point estimate relative to DiD estimators. Filled points are post-treatment periods.

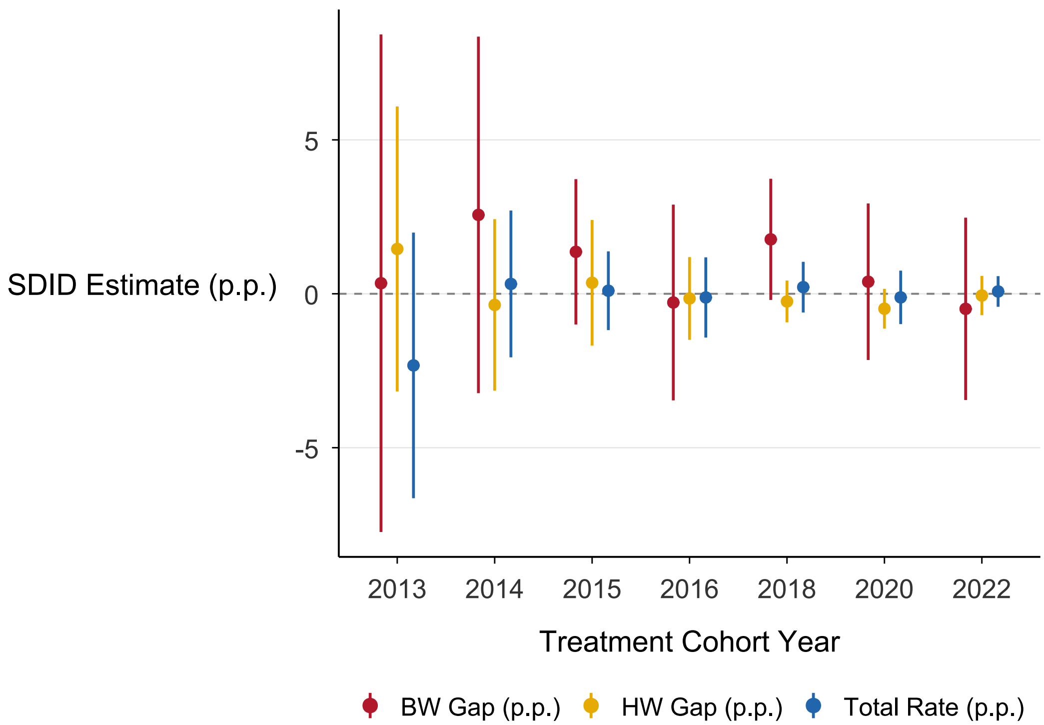


Figure A20: SDID Estimates by Treatment Cohort

Notes: Cohort-specific SDID ATT estimates with 95% bootstrap confidence intervals for BW gap, HW gap, and total suspension rate. Most cohort-specific estimates for the BW gap are positive, though only the largest cohorts reach conventional significance. HW gap and total rate estimates are centered near zero.

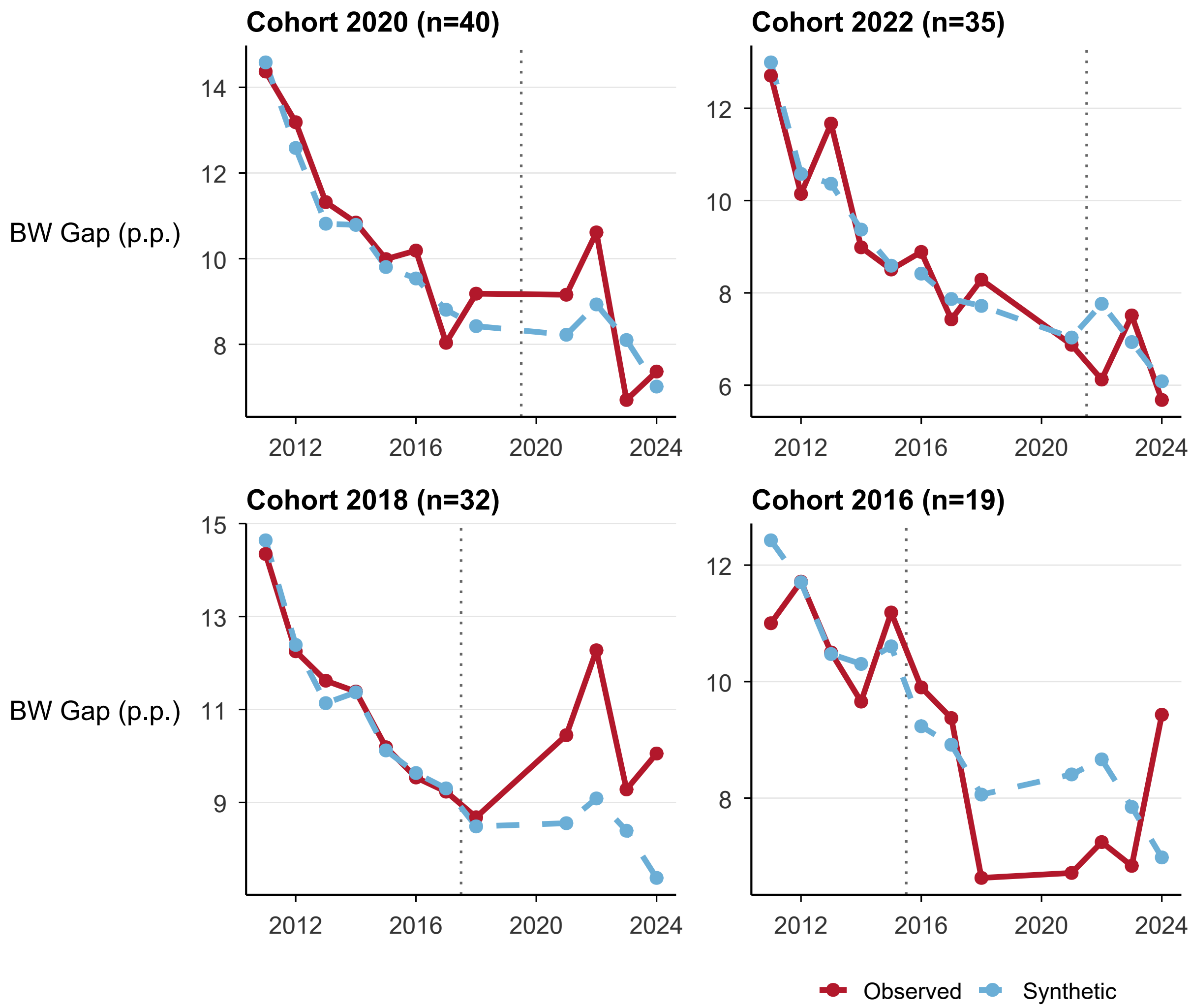


Figure A21: SDID Counterfactual Trajectories: BW Suspension Gap

Notes: Observed treated-group mean vs. SDID synthetic control for the four largest treatment cohorts. Each panel shows the outcome trajectory in levels. The dotted vertical line marks the cohort's first treatment year. Following Lal and Thompson (2024), no error bars are shown on the trajectory plot.

D Robustness Details

This appendix provides additional detail on the robustness checks summarized in Section 4.4. Across nine alternative specifications, the Sun-Abraham ATT for the BW gap ranges from +2.3 to +3.9 percentage points, always statistically significant at the 5% level (Figure 7 in the main text). Figure A22 reports the Rambachan and Roth (2023) sensitivity analysis under relative magnitudes restrictions, showing that the BW gap confidence interval includes zero when post-treatment trend violations are allowed to exceed the largest pre-treatment violation. Figure A23 presents the randomization inference distribution (RI $p = .004$). Figures A24–A29 report placebo outcome tests, the CS not-yet-treated comparison, leave-one-cohort-out stability for the HW gap, and the full specification curve.

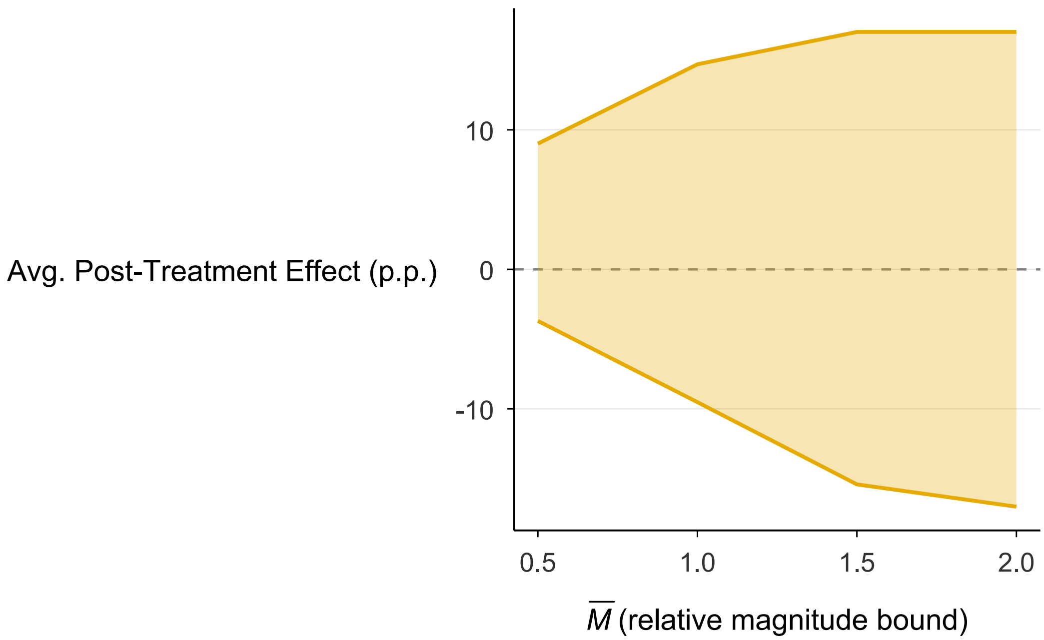


Figure A22: HonestDiD: Relative Magnitudes Sensitivity
Notes: Robust confidence sets from Rambachan and Roth (2023) under relative magnitudes restrictions for the BW gap effect. The x-axis parameterizes the maximum ratio of the post-treatment violation of parallel trends to the largest pre-treatment violation.

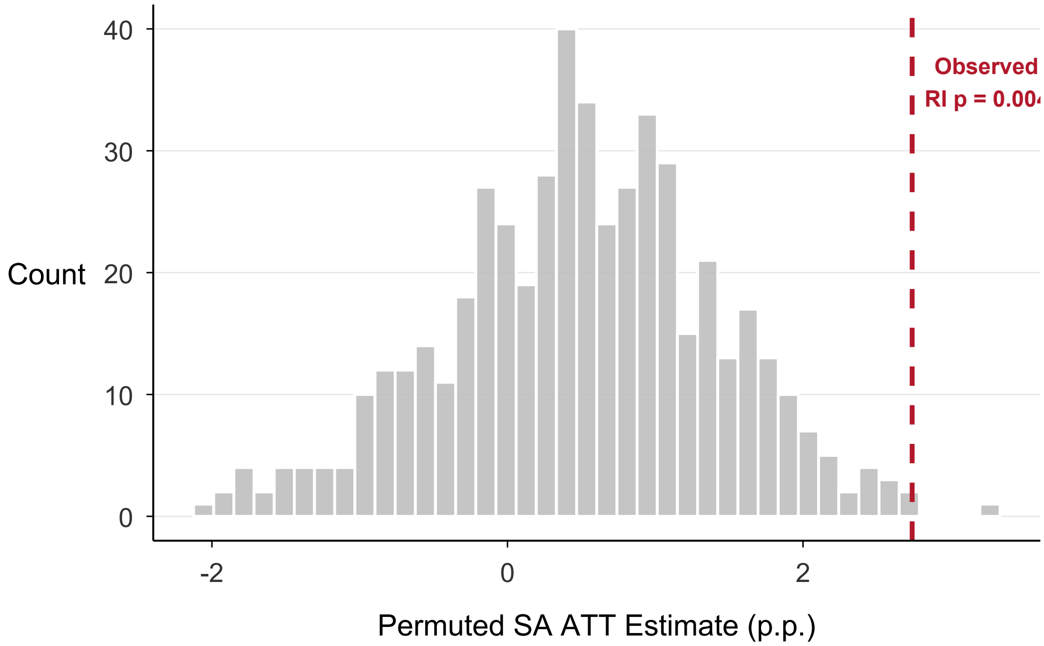


Figure A23: Randomization Inference: BW Gap

Notes: Distribution of placebo Sun-Abraham ATT estimates from 500 random permutations of treatment timing. The vertical red line marks the observed estimate. The randomization inference p-value is 0.004 (2 of 500 permutations exceed the observed value).

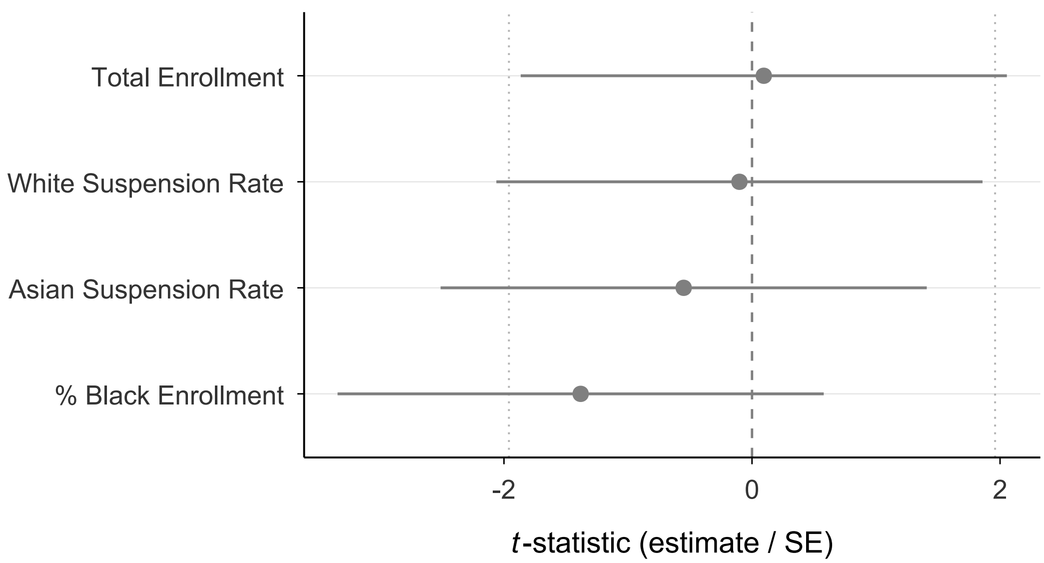


Figure A24: Placebo Outcome Tests

Notes: Sun-Abraham ATT t-statistics (estimate/SE) for placebo outcomes that should not be affected by SMD adoption. Dotted lines mark the ± 1.96 critical values. All estimates are small and statistically insignificant.

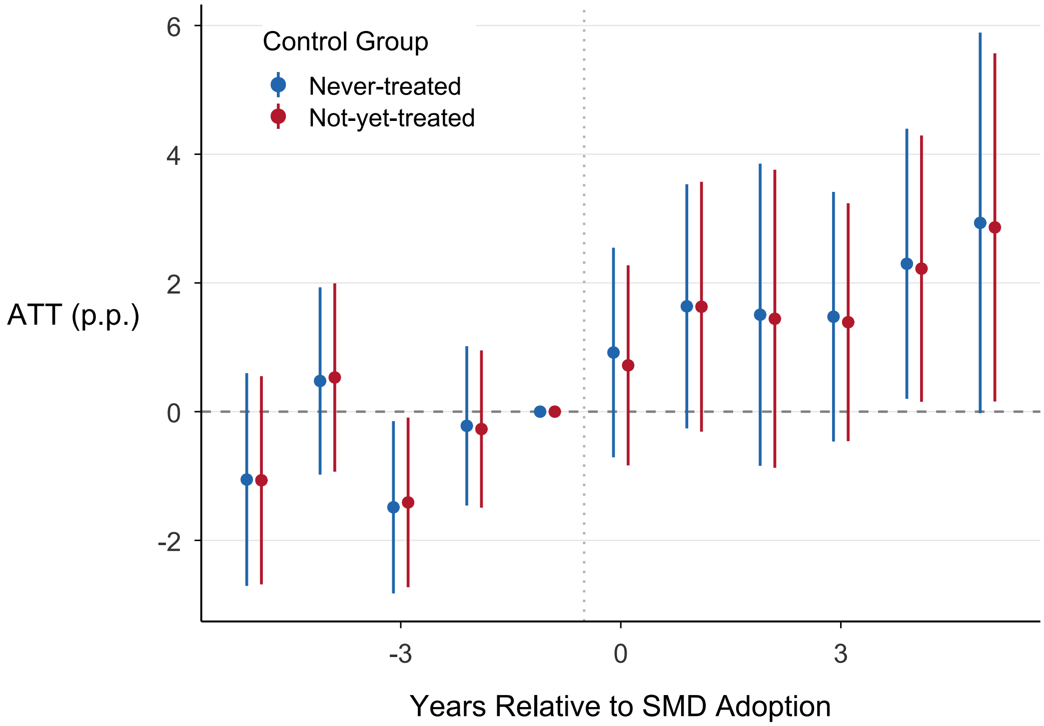


Figure A25: Callaway-Sant'Anna Event Study: Not-Yet-Treated Comparison
Notes: Callaway-Sant'Anna event study using only not-yet-treated units as the comparison group, rather than never-treated units. Results are similar to the baseline CS specification.

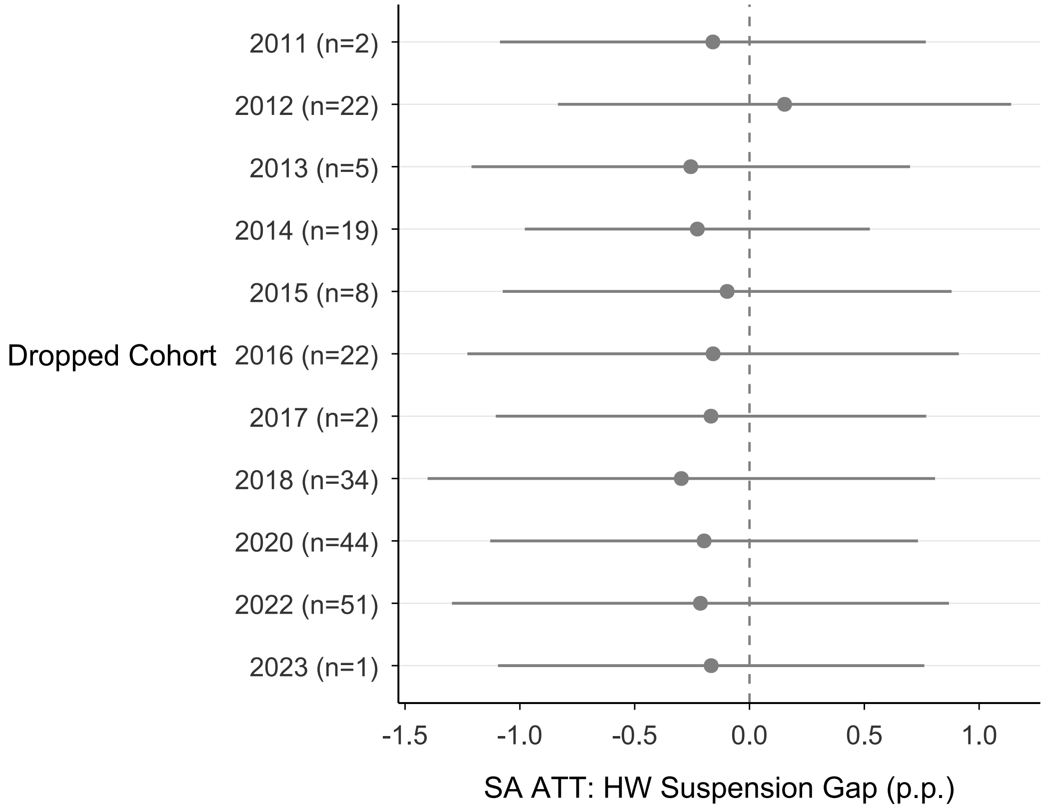


Figure A26: Leave-One-Cohort-Out: HW Suspension Gap

Notes: SA ATT estimates for the HW suspension gap when each treatment cohort is excluded. All estimates remain close to zero and statistically insignificant, confirming the null result for the HW gap.

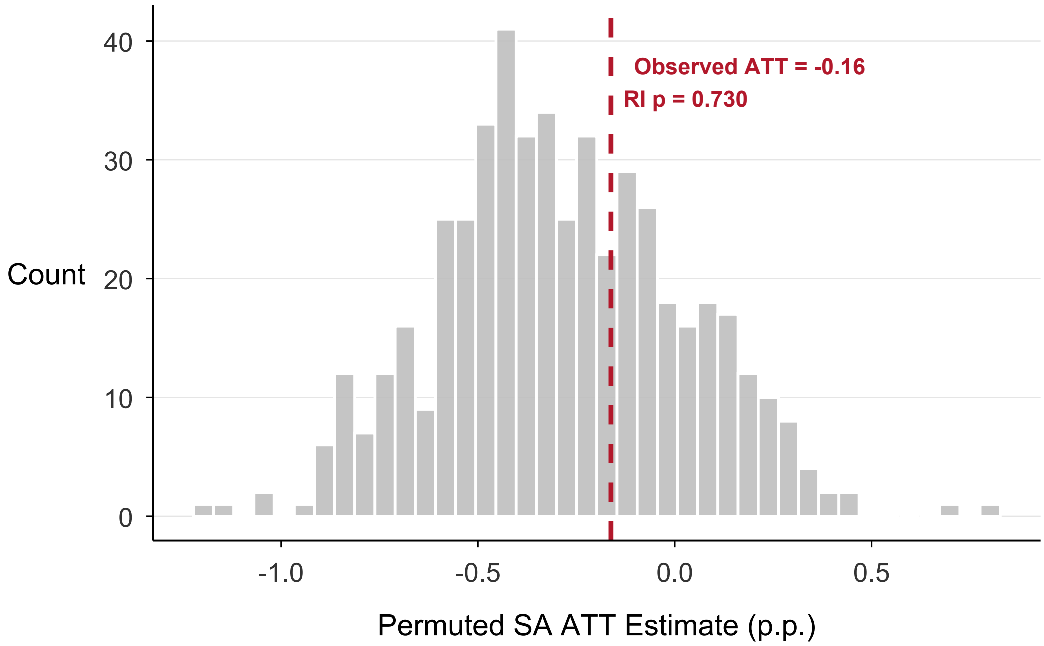


Figure A27: Randomization Inference: HW Gap

Notes: Distribution of placebo SA ATT estimates for the HW gap from 500 random permutations of treatment timing. The observed estimate falls well within the placebo distribution ($p = .73$), consistent with a null effect.

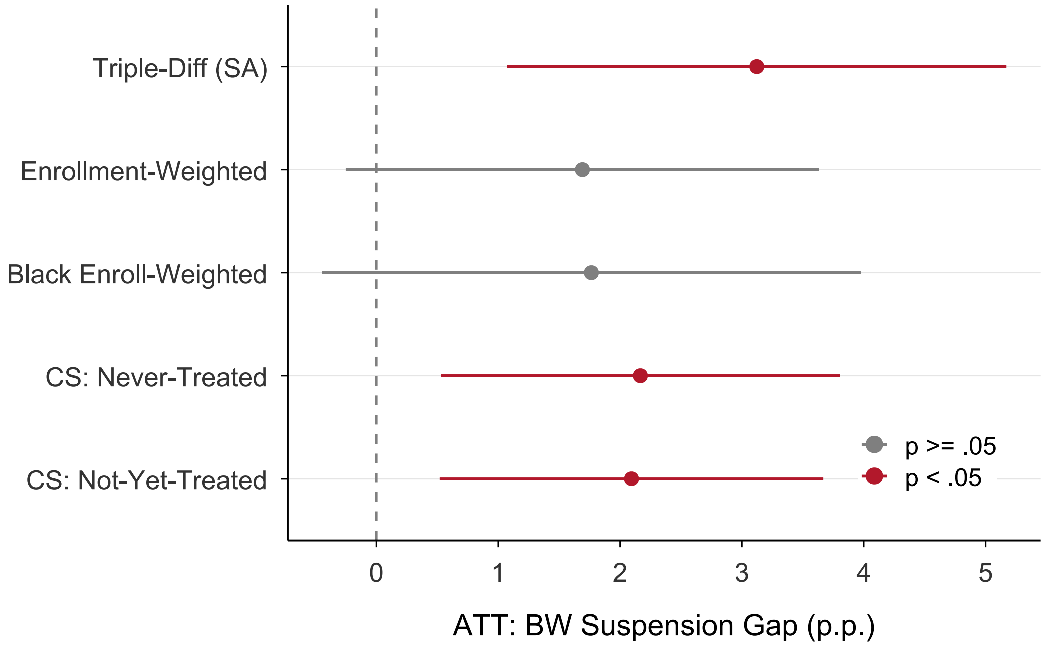


Figure A28: Additional Robustness Checks: Summary

Notes: Summary of additional robustness checks including triple-difference, enrollment weighting, CS not-yet-treated control group, and placebo outcomes. The BW gap result is robust across all specifications.

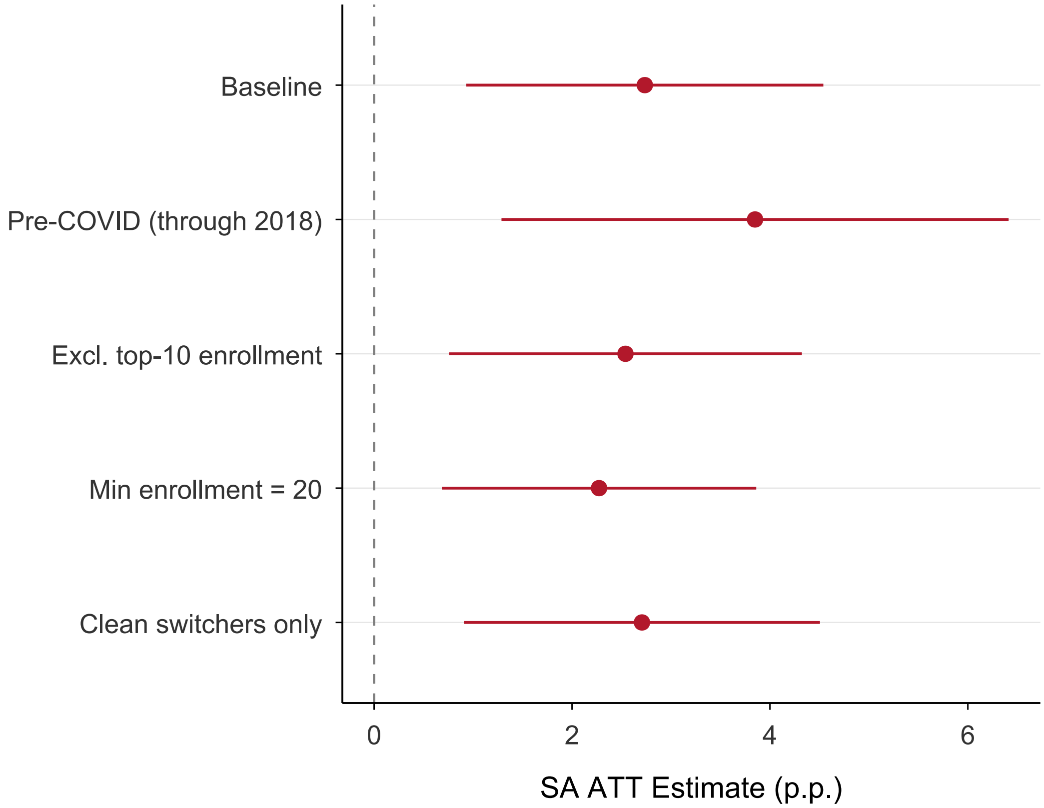


Figure A29: Robustness: SA BW Gap Across All Specifications

Notes: Sun-Abraham ATT estimates for the BW suspension gap across all robustness specifications including the baseline, pre-COVID only, excluding top-10% enrollment, minimum enrollment 20, clean switchers, and log/asinh transformations. Points are estimates; whiskers show 95% confidence intervals.

E Heterogeneity

This appendix presents the full subgroup heterogeneity results summarized in Section 4. Figure A30 reports SA ATT estimates for the BW suspension gap across subgroups defined by district size, minority student share, Black student share, Hispanic student share, and pre-treatment gap level. The BW gap effect is notably larger in districts with low Black student shares (+10.0 pp vs. +1.8 pp in high-share districts), consistent with small-denominator rate volatility (districts with few Black students exhibit larger swings in race-specific suspension rates) and threshold-driven sample selection (the minimum-enrollment criterion selects for districts where small fluctuations produce large rate changes). Figure A31 shows the corresponding HW gap estimates, which are uniformly null. Figure A32 disaggregates by board composition gains after SMD adoption, showing that districts gaining a predicted-Black board member exhibit a similar BW gap increase to those gaining only Hispanic members.

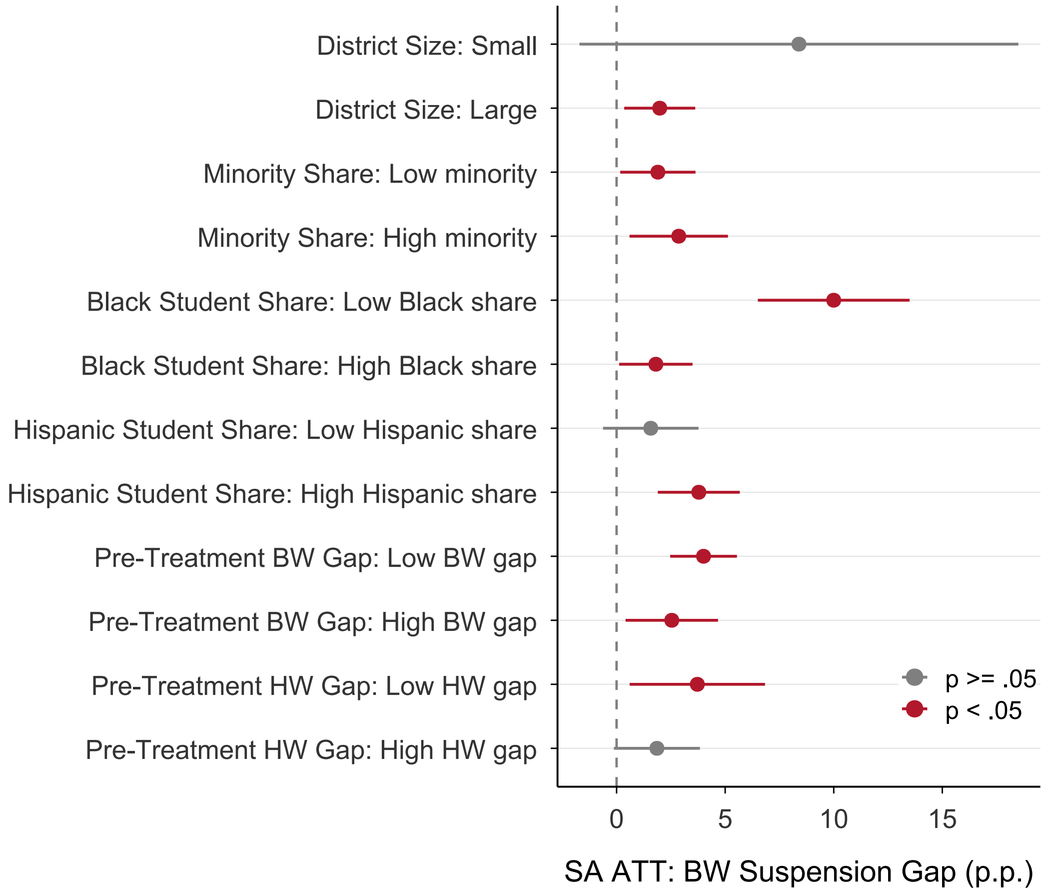


Figure A30: Subgroup Heterogeneity: BW Suspension Gap

Notes: Forest plot of Sun-Abraham ATT estimates for the BW suspension gap across subgroups defined by district characteristics. Points are estimates; whiskers show 95% confidence intervals. The BW gap effect is larger in districts with low Black student shares (+10.0 vs. +1.8 in high-share districts) and smaller in large districts.

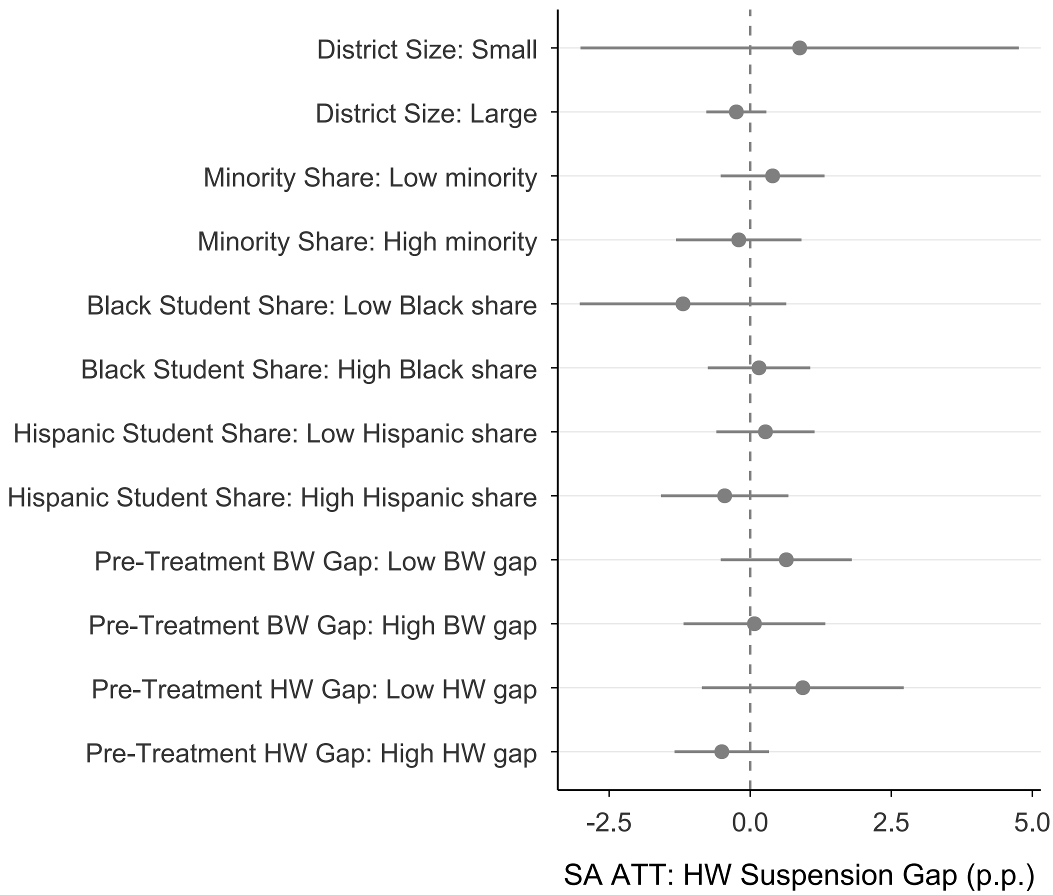


Figure A31: Subgroup Heterogeneity: HW Suspension Gap

Notes: Forest plot of Sun-Abraham ATT estimates for the HW suspension gap across subgroups. All estimates are close to zero and insignificant.

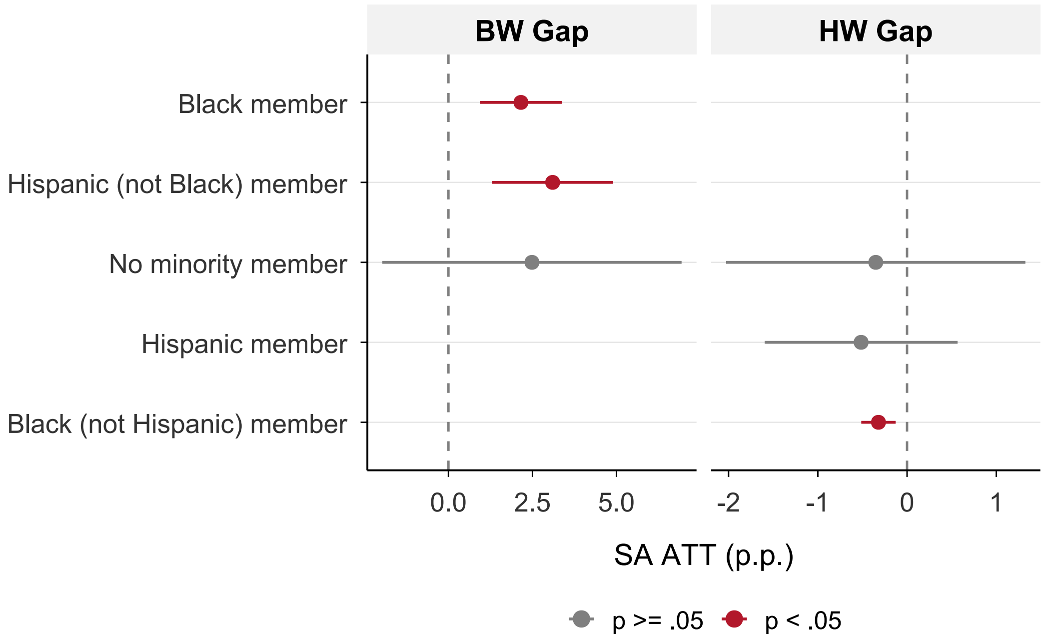


Figure A32: Heterogeneity by Board Racial Composition Gains

Notes: Sun-Abraham ATT estimates separately for districts that gained a predicted-Black board member after SMD adoption (17 districts), gained a predicted-Hispanic member but not Black (140 districts), and gained no minority member (40 districts), among treated districts with valid BW gap data and post-treatment board elections. Consistent with the race-match hypothesis, the BW gap increase is larger in districts without Black representation gains.

F Teacher Diversity Analysis

This appendix examines whether SMD adoption affects the racial composition of the teaching workforce and whether teacher diversity mediates the discipline gap effects. I construct a harmonized teacher demographics panel from CDE's StaffDemo files (individual-level, 2011–2017) and Staff by Race/Ethnicity files (aggregate, 2021–2024), covering 722 of 788 analysis districts. Figure A33 plots minority teacher share trends for treated and control districts. Figure A34 presents event study estimates for the teacher diversity first stage, showing gradual increases in minority teacher shares after SMD adoption. Figure A35 summarizes the first-stage estimates: SMD adoption increases minority teacher share by approximately

1.5 percentage points (SA, $p = .03$), driven by Hispanic teachers. Figure A36 reports the key mediation result: Black teacher share strongly moderates the BW discipline gap effect (-82 to -96 points, $p < .001$), while Hispanic teacher share does not—a pattern consistent with the race-match hypothesis and parallel to the board composition findings in Section 5.

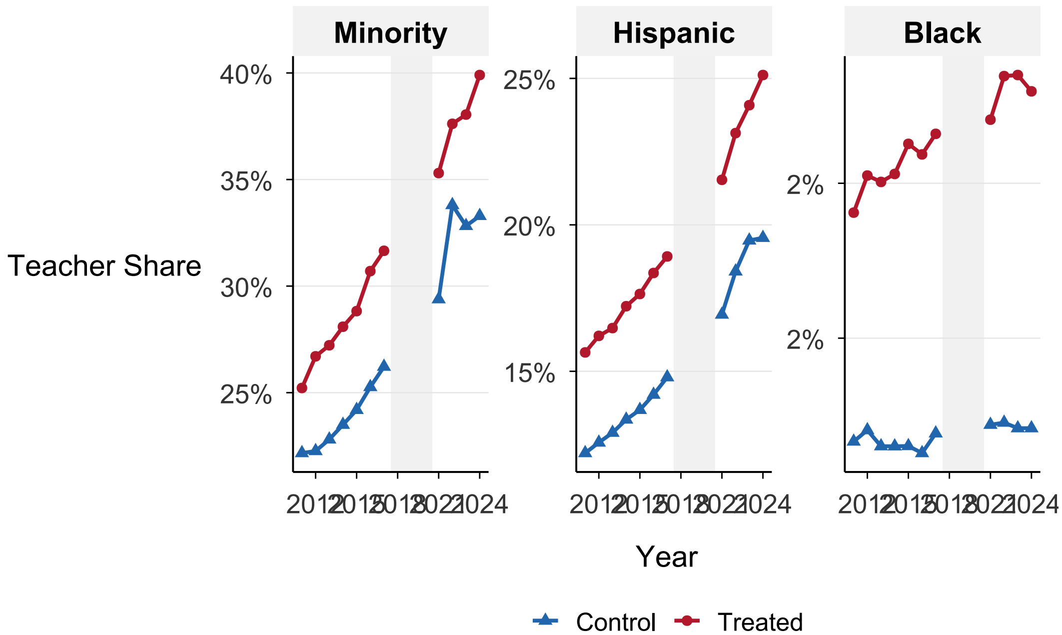


Figure A33: Teacher Workforce Diversity Trends

Notes: Mean teacher racial shares over time, separately for treated and control districts. Both groups show gradual increases in minority teacher shares, with treated districts experiencing slightly faster growth after SMD adoption.

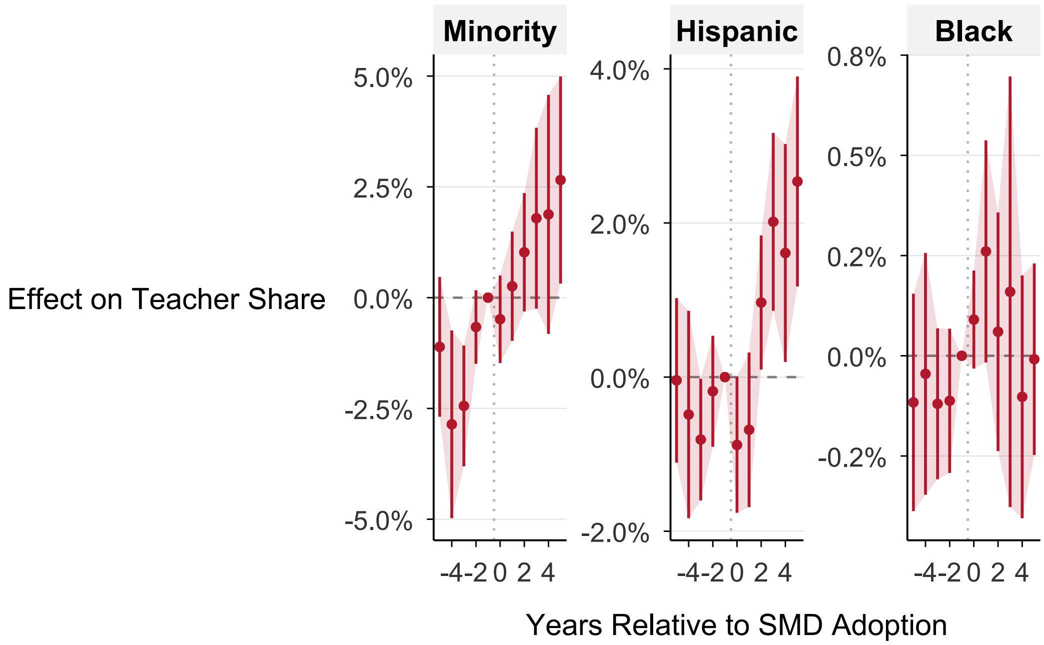


Figure A34: Event Study: Teacher Diversity First Stage

Notes: Event study estimates for the effect of SMD adoption on minority, Hispanic, and Black teacher shares. Effects emerge gradually after treatment, consistent with hiring being a slow-moving process influenced by board priorities.

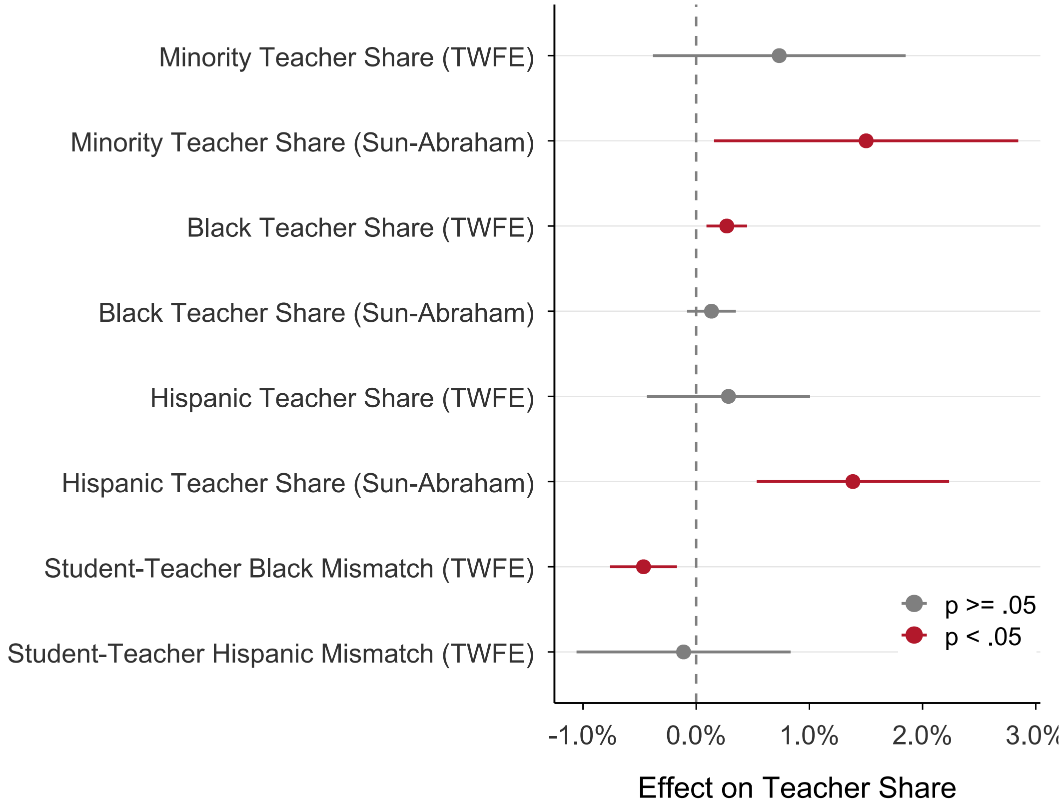


Figure A35: Teacher Diversity First Stage: Summary

Notes: Forest plot of TWFE and Sun-Abraham estimates for the effect of SMD adoption on various teacher diversity measures. Minority teacher share: SA +1.5 pp ($p = .029$); Hispanic teacher share: SA +1.4 pp ($p = .001$); Black teacher share: TWFE +0.3 pp ($p = .003$).

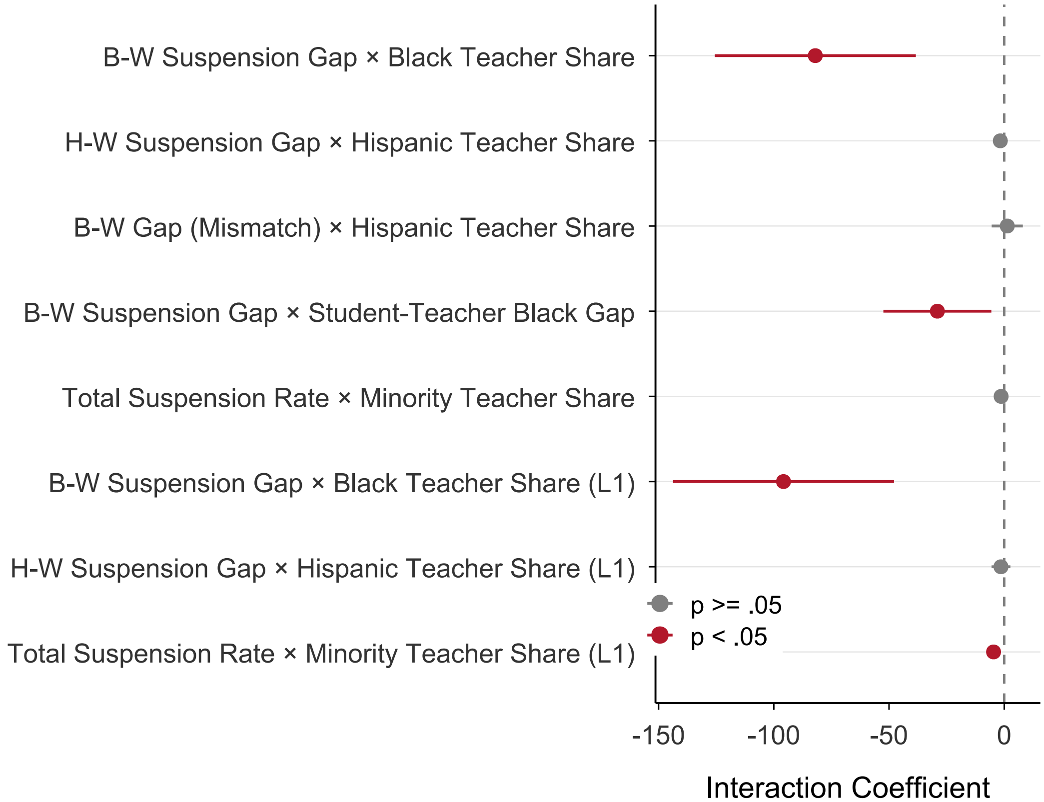


Figure A36: Teacher Diversity as Mediator of Discipline Effects

Notes: Interaction coefficients from models interacting treat_post with teacher racial shares. Black teacher share strongly moderates the BW gap effect (-82 to -96 points, $p < .001$), while Hispanic teacher share does not (mismatch test: $+1.3$, $p = .70$). The student-teacher Black mismatch gap also significantly moderates the BW effect (-29, $p = .016$).

G Spending Channel Details

This appendix investigates whether per-pupil spending changes after SMD adoption and whether such changes mediate the discipline gap effects. Figure A37 presents the spending event study, showing a modest increase in per-pupil expenditure after adoption—consistent with Fischer (2023). Figure A38 plots raw spending trends for treated and control districts. Figure A39 reports the spending moderation test: the interaction of SMD treatment with log per-pupil spending is not significant for either the BW or HW gap, indicating that spending changes do not drive the discipline effects. Figure A40 provides context by comparing the BISG-predicted racial composition of elected board members before and after SMD adoption.

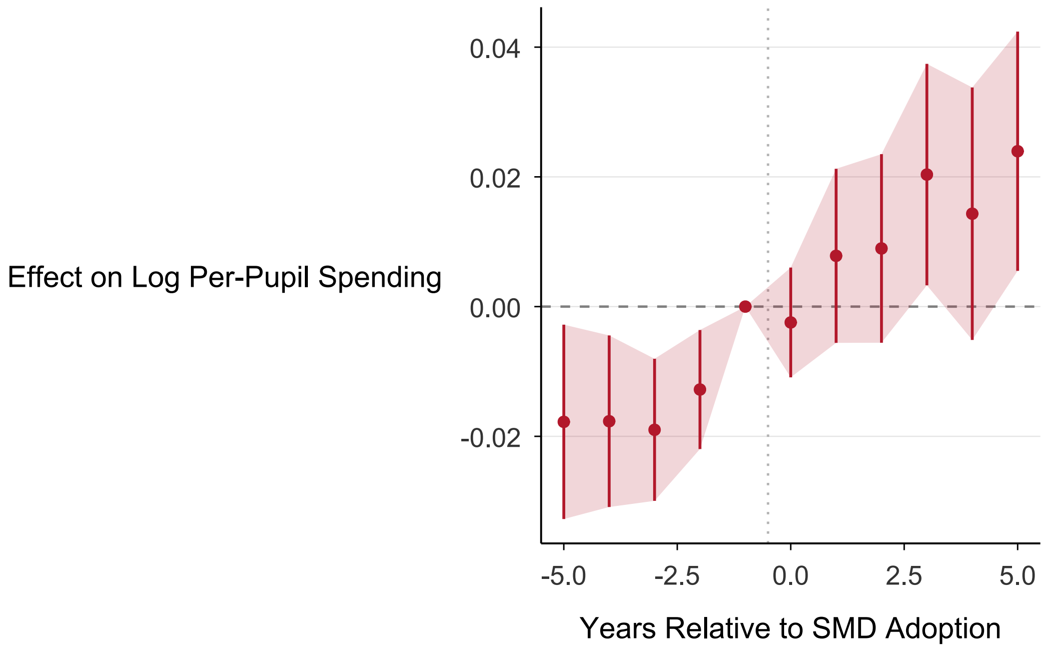


Figure A37: Event Study: Per-Pupil Spending

Notes: Event study estimates for the effect of SMD adoption on log per-pupil expenditure. Spending increases modestly after adoption, consistent with Fischer (2023). However, spending does not moderate the BW discipline gap effect (Section 5.4).

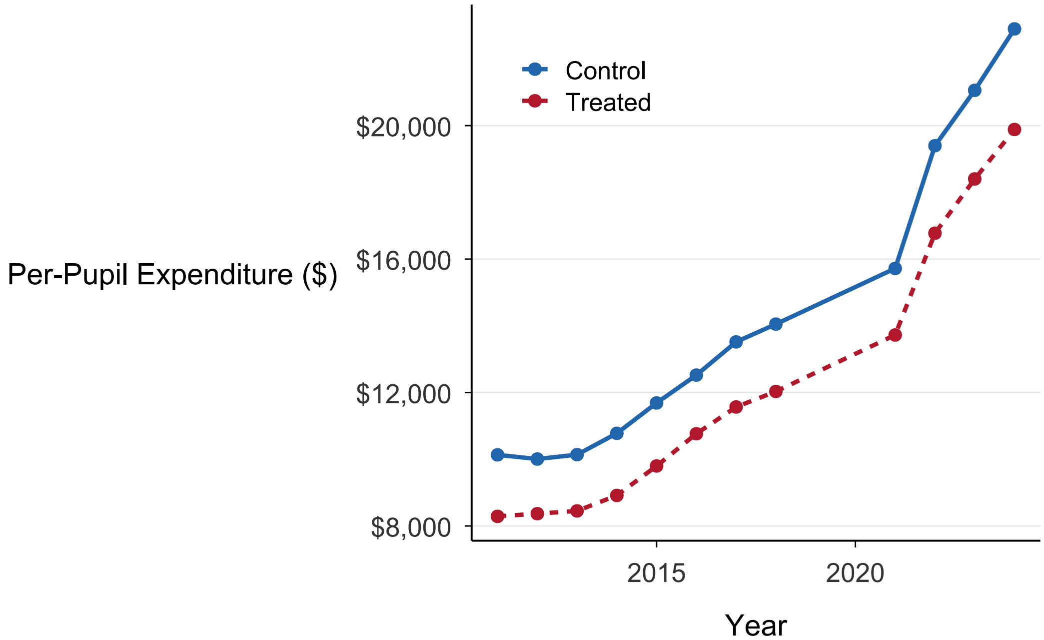


Figure A38: Per-Pupil Spending Trends by Treatment Group

Notes: Mean per-pupil expenditure over time for treated and control districts.

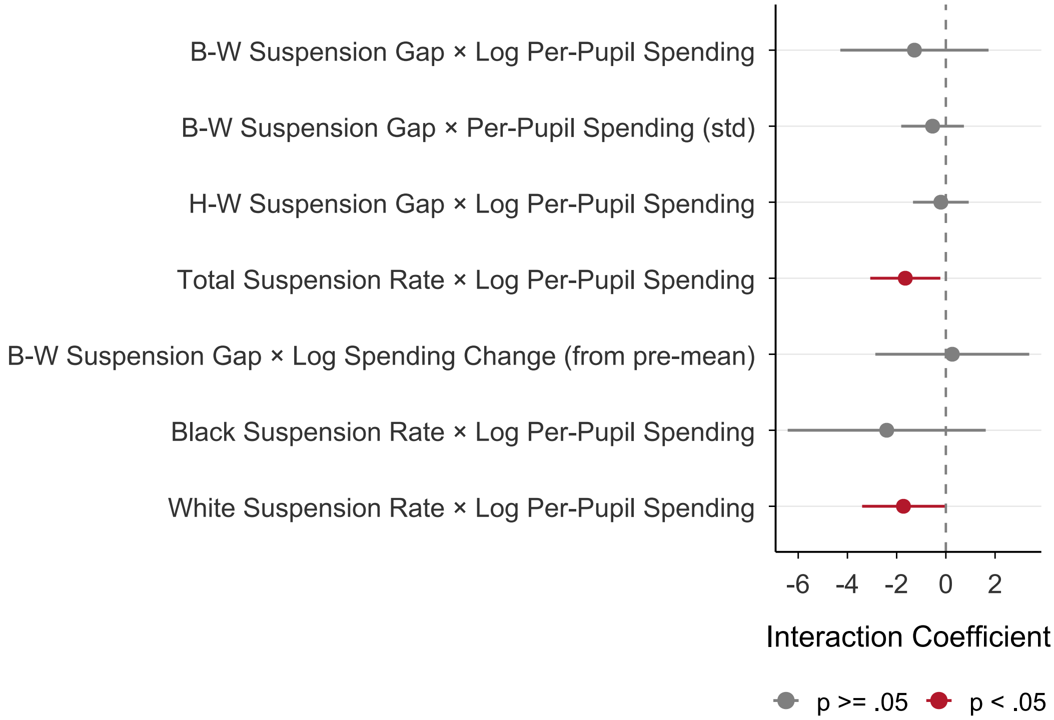


Figure A39: Spending as Moderator of Discipline Effects

Notes: Interaction of SMD treatment with log per-pupil spending on the BW and HW suspension gaps. Neither interaction is significant, indicating that spending does not moderate the discipline gap effects.

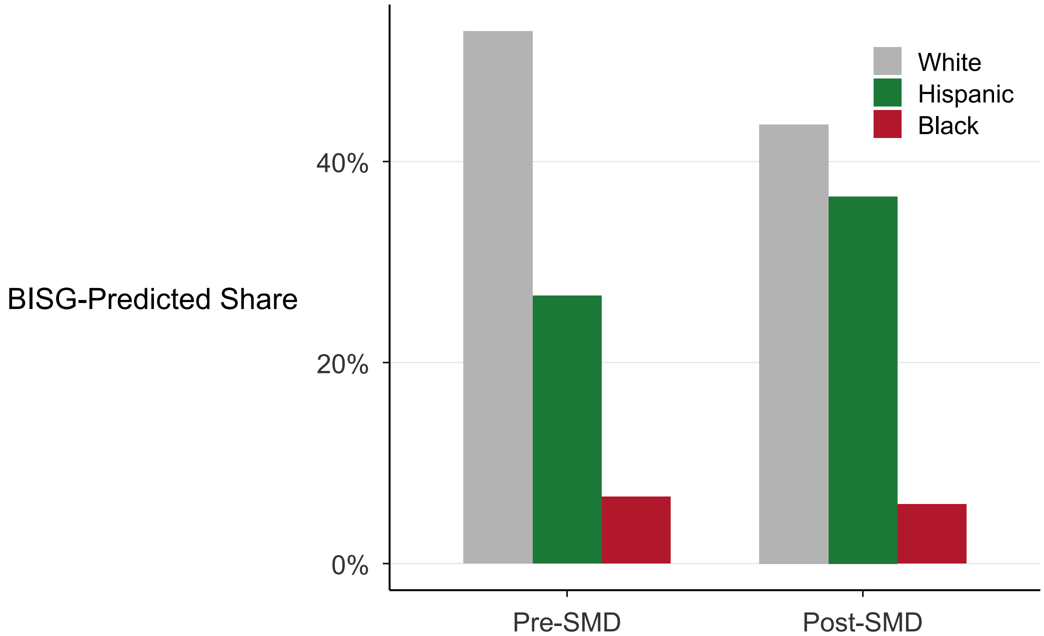


Figure A40: Board Composition: Pre vs. Post SMD Adoption
Notes: Distribution of BISG-predicted racial composition of elected board members before and after SMD adoption, across all treated districts. Hispanic share increases substantially; Black share remains essentially unchanged.

H Achievement Gap Outcomes (SEDA 6.0)

As a domain-specific falsification check, I test whether SMD adoption affects standardized test score gaps using the Stanford Education Data Archive (SEDA) version 6.0 (Reardon et al. 2026). Unlike the placebo outcomes in Section 4.4, achievement is a domain where governance changes could plausibly have effects—if SMD elections improve broad district quality, we would expect improvements in academic outcomes as well. A null on achievement is therefore informative but not guaranteed, and this analysis is properly framed as a domain-specific falsification check rather than a placebo test.

Data. SEDA 6.0 provides district-level, grade-level, subject-specific mean test scores and racial achievement gaps in cohort-scale standard deviation (SD) units, covering 2009–2018 (with 2014 missing due to the CST-to-SBAC transition in California). I aggregate to the district-year level by weighting across grades and subjects by assessment count, then merge with the analysis panel via a CDE–NCES district crosswalk. The resulting sample covers approximately 640 districts (198 treated, 442 control) observed over 8 years, yielding approximately 4,400 district-year observations for overall achievement, 2,800 for the W-H gap, and 1,100 for the W-B gap (which requires sufficient Black enrollment). Treatment does not predict SEDA data availability ($p = .62$), confirming that the merged sample is not selected on treatment.

Results. Table A12 reports the main estimates. Across all three estimators (TWFE, Sun-Abraham, Callaway-Sant'Anna), the effects on achievement gaps are null. The preferred SA estimates are +0.012 SD (SE = 0.012; $p = .33$) for the W-H gap and +0.004 SD (SE = 0.014; $p = .78$) for the W-B gap. The CS estimates, which use a balanced subsample restricted to in-range cohorts, are similarly null. TWFE shows a significant negative effect on overall achievement levels (−0.046 SD; $p < .001$), but this vanishes under SA (−0.012 SD; $p = .44$), reflecting classic staggered-adoption bias in the TWFE estimator.

Table A13 provides the key falsification test: a side-by-side comparison of achievement and discipline effects on the same SEDA-available district×year cells. While achievement gaps are uniformly null, the discipline BW suspension gap remains significant (+3.29 pp; SE = 1.32; $p = .01$) on this restricted sample. The discipline HW gap is also significant on this subsample (−1.28 pp; SE = 0.60; $p = .03$), in contrast to the full-sample null. This subsample difference likely reflects the SEDA-available districts being larger and more urban, with higher Hispanic enrollment shares and longer post-treatment exposure windows; the full-sample null remains the more conservative and appropriate estimate for the HW gap. The key finding is that discipline effects are domain-specific rather than artifacts of sample construction.

Figure A41 presents achievement gap trends and event studies. Pre-treatment trends are parallel for both the W-H and W-B gaps, and the event study coefficients are flat both before and after treatment. Pre-trend Wald tests are clean ($p = 0.58$ for both gaps).

Table A12: Domain-Specific Falsification: Achievement Gap Outcomes (SEDA 6.0)

Outcome / Estimator	ATT	(SE)	N
W-H Achievement Gap			
TWFE	0.006	(0.010)	2,771
Sun-Abraham	0.012	(0.012)	2,771
CS	0.023	(0.014)	1,820
W-B Achievement Gap			
TWFE	0.007	(0.015)	1,138
Sun-Abraham	0.004	(0.014)	1,138
CS	−0.011	(0.016)	637
TWFE [BW-eligible]	0.010	(0.015)	1,100
Sun-Abraham [BW-eligible]	0.005	(0.014)	1,100
CS [BW-eligible]	−0.011	(0.016)	637
Overall Achievement			
TWFE	-0.046^{***}	(0.012)	4,407
Sun-Abraham	−0.012	(0.015)	4,407
CS	−0.033	(0.017)	3,423

Notes: ATT estimates in cohort-scale standard deviation units (SEDA 6.0). SEDA data cover 2009–2018 (no 2014 due to CST→SBAC transition). CS estimates use a balanced panel restricted to cohorts within the data year range (≥3 districts per cohort). BW-eligible restricts to districts with mean pre-treatment Black enrollment ≥20. SEs clustered at district level. $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure A42 displays the achievement vs. discipline comparison as a forest plot, making the contrast between null achievement effects and significant discipline effects visually immediate.

Table A14 reports robustness checks including an alternative reference period ($t = -2$), grade-level panel with district×grade fixed effects, separate estimates for math and reading, and SEDA SE-based filtering. All gap estimates remain null across specifications. The final rows report minimum detectable effects (MDEs) at 80% power: 0.034 SD for the W-H gap and 0.040 SD for the W-B gap, both well below the discipline-proportional benchmarks (0.055 SD), confirming adequate statistical power.

Table A13: Achievement vs. Discipline: Same-Sample Comparison (Sun-Abraham)

Outcome	ATT	(SE)	N
Panel A: Achievement (SD units)			
W-H Achievement Gap	0.012	(0.012)	2,771
Overall Achievement	−0.012	(0.015)	4,407
W-B Achievement Gap	0.004	(0.014)	1,138
Panel B: Discipline (p.p., same sample)			
B-W Suspension Gap	3.29^{**}	(1.32)	2,896
H-W Suspension Gap	-1.28^{**}	(0.60)	4,276
Total Suspension Rate	−0.22	(0.45)	4,375

Notes: Sun-Abraham ATT estimates. Panel A: SEDA 6.0 achievement outcomes (SD units). Panel B: CDE suspension outcomes (percentage points) estimated on the same district×year cells where SEDA data are available. SEs clustered at district level. $^{**}p < 0.05$, $^{***}p < 0.01$.

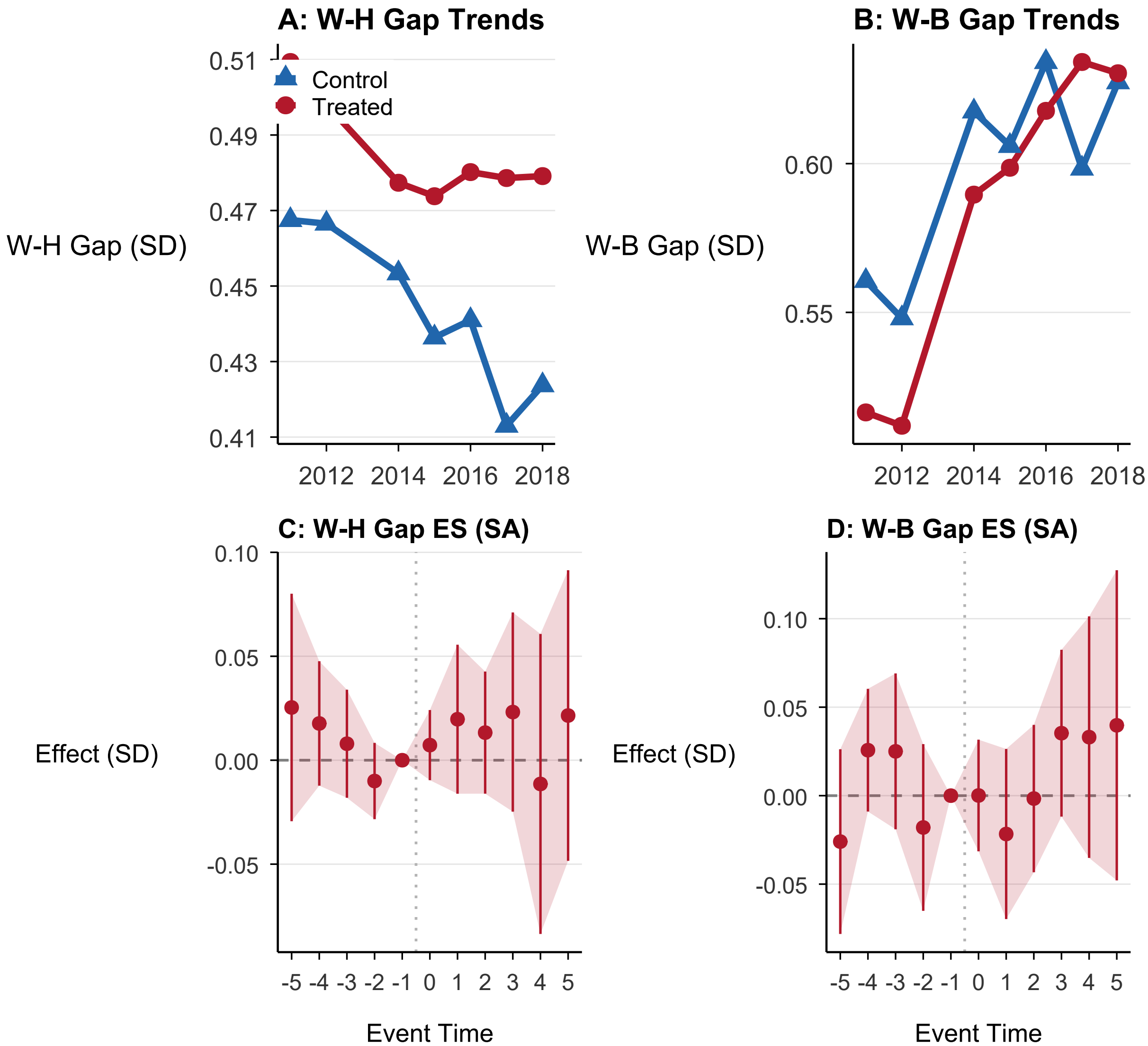


Figure A41: Achievement Gap Trends and Event Studies (SEDA 6.0)

Notes: Panels A–B show mean W-H and W-B standardized test score gaps (cohort-scale SD units) from SEDA 6.0, 2009–2018. Panels C–D show Sun-Abraham event study coefficients with 95% CIs; the omitted period is $t-1$. All post-treatment coefficients are near zero, consistent with no effect of SMD adoption on achievement gaps.

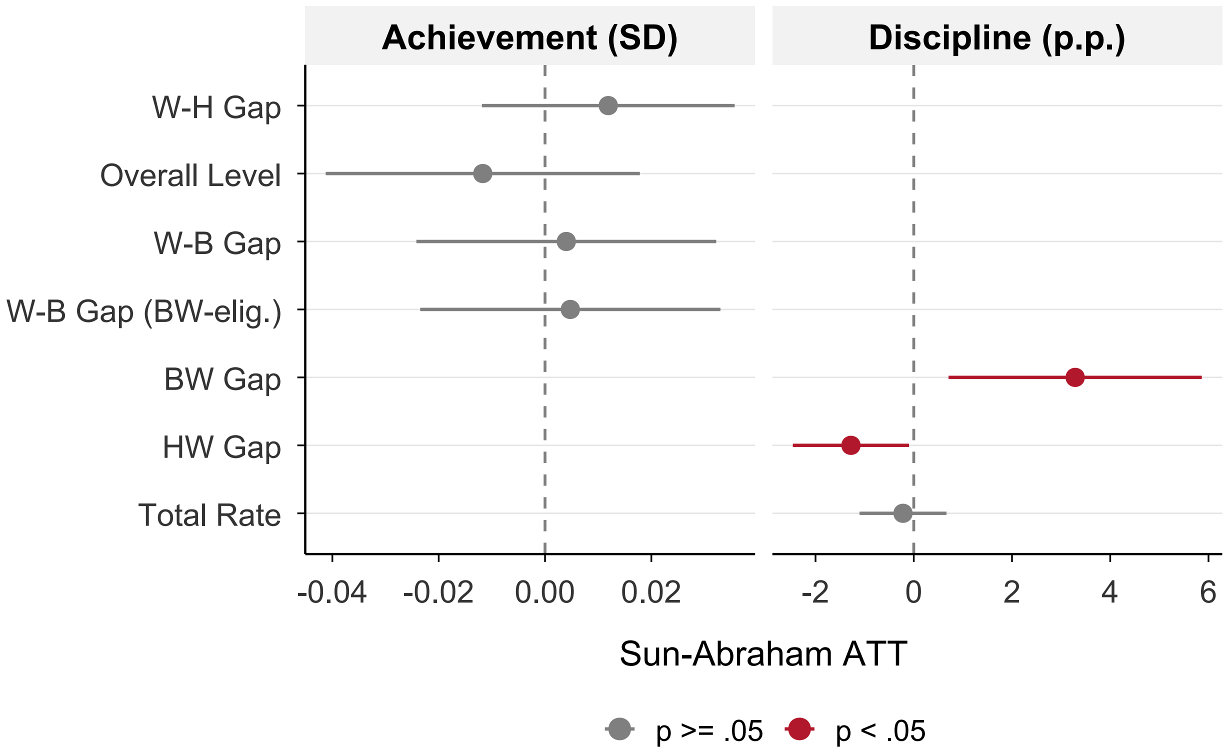


Figure A42: Achievement vs. Discipline: Same-Sample Comparison

Notes: Sun-Abraham ATT estimates with 95% confidence intervals. Left panel: SEDA achievement outcomes (SD units). Right panel: CDE discipline outcomes (percentage points) estimated on the same district×year cells where SEDA data are available. Achievement effects are uniformly null; discipline effects remain significant.

Table A14: Achievement Gap Robustness and Power

	W-H Gap		W-B Gap	
Specification	ATT	(SE)	ATT	(SE)
SA ref.p=-2	0.000	(0.009)	0.015	(0.022)
Grade-level TWFE	0.004	(0.008)	—	—
Grade-level SA	0.001	(0.008)	—	—
Math TWFE	0.004	(0.012)	0.003	(0.017)
Math SA	0.011	(0.015)	0.007	(0.017)
Reading TWFE	0.010	(0.009)	0.003	(0.015)
Reading SA	0.015	(0.012)	0.004	(0.016)
SE<=p75 TWFE	0.006	(0.007)	−0.003	(0.011)
SE<=p75 SA	0.013	(0.010)	−0.007	(0.013)
MDE (80% power)	0.034		0.040	
Disc. proportional benchmark	0.055		0.055	

Notes: Robustness checks for achievement gap outcomes. All estimates in cohort-scale SD units. Grade-level specifications use district×grade fixed effects. SE≤p75 drops observations with SEDA standard errors above the 75th percentile. MDE at 80% power = 2.802× SE. Discipline proportional benchmark = 0.225 × outcome SD. $^{**}p < 0.05$, $^{***}p < 0.01$.