

The Dog That Didn't Bark: Educational Content Restriction Laws and Teacher Labor Markets*

APEP Autonomous Research[†] @SocialCatalystLab

February 11, 2026

Abstract

Between 2021 and 2023, twenty-three U.S. states enacted laws restricting classroom instruction on race, gender, and “divisive concepts,” prompting claims of an impending teacher exodus. I test this prediction using Census Quarterly Workforce Indicators for K–12 schools (NAICS 6111) in a staggered difference-in-differences design with the Callaway and Sant’Anna (2021) estimator. The overall ATT for log employment is 0.023 (SE = 0.020), statistically indistinguishable from zero. Separations, hiring, earnings, and female workforce share are similarly null. The conventional TWFE estimator spuriously finds a significant positive effect (0.109, $p < 0.05$), illustrating how naive two-way fixed effects can produce misleading results under staggered adoption. One margin responds: turnover rises significantly (0.0048, $p < 0.05$), suggesting increased churn without net employment loss. The design can detect effects of 5.5% or larger with 80% power.

JEL Codes: J45, I28, J63, K31

Keywords: divisive concepts, teacher labor markets, content restriction laws, difference-in-differences, Callaway-Sant’Anna, null result

*This paper is a revision of APEP-0222. See [https://github.com/SocialCatalystLab/ape-papers/
tree/main/apep_0222](https://github.com/SocialCatalystLab/ape-papers/tree/main/apep_0222) for the original.

[†]Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch

1. Introduction

By the end of 2023, twenty-three U.S. states had passed laws restricting how teachers can discuss race, gender, and history. Headlines in the *New York Times* and *Washington Post* warned of an impending “teacher exodus,” fueled by accounts of educators abandoning their classrooms in the face of new legal threats (Kraft and Lyon, 2023; García and Weiss, 2022). National Education Association surveys reported that large majorities of teachers felt increased stress from legislative scrutiny of curricula (National Education Association, 2022). If these laws are making the profession untenable, the effects should be visible in the data.

Yet the empirical basis for these claims is remarkably thin. The existing evidence consists almost entirely of qualitative surveys, journalistic accounts, and cross-sectional correlations. No study has applied modern causal inference methods to estimate the effect of content restriction laws on teacher labor market outcomes. This paper fills that gap. If these laws are indeed causing a teacher exodus, it should be detectable in administrative employment data: we would observe increased separations, reduced hiring, falling employment, and potentially rising wages as districts compete for a shrinking pool of educators.

I use quarterly administrative data from the Census Bureau’s Quarterly Workforce Indicators (QWI) covering all fifty states and the District of Columbia from 2015 through 2024. A key innovation of this revision is the use of four-digit NAICS industry data: I focus on NAICS 6111 (Elementary and Secondary Schools) rather than the broader NAICS 61 (Educational Services) used in earlier work on this topic. NAICS 6111 isolates precisely the K–12 school workforce that content restriction laws target, avoiding dilution from universities, tutoring companies, and other educational services that are largely unaffected by these statutes. The QWI provides state-quarter level counts of employment, hires, separations, and average earnings derived from near-universal state unemployment insurance records. I exploit the staggered adoption of content restriction laws across twenty-three states between 2021 and 2023 in a difference-in-differences framework, using the twenty-eight state-equivalents (including DC) that never enacted such legislation as controls.

The primary estimator is the group-time average treatment effect of Callaway and Sant’Anna (2021), which avoids the well-documented biases of conventional two-way fixed effects (TWFE) estimation under heterogeneous treatment timing (Goodman-Bacon, 2021; de Chaisemartin and D’Haultfoeuille, 2020; Sun and Abraham, 2021; Borusyak et al., 2024). I aggregate group-time effects to an overall ATT and to dynamic event-study coefficients. Robustness checks include the Sun and Abraham (2021) interaction-weighted estimator, a triple-difference design comparing K–12 schools to healthcare, placebo sector tests, heterogeneity by law stringency, randomization inference, minimum detectable effect analysis, and

a comparison of NAICS 6111 versus the broader NAICS 61 sector.

The results are clear and consistent: content restriction laws have no detectable effect on aggregate K–12 teacher labor market outcomes. The Callaway-Sant’Anna ATT for log employment is 0.023 (SE = 0.020), a 2.3% point estimate that is statistically indistinguishable from zero. Effects on separation rates (0.003, SE = 0.007), log earnings (−0.013, SE = 0.010), and hire rates (0.006, SE = 0.009) are similarly small and statistically insignificant. The null persists across every alternative specification. Event-study coefficients show clean pre-trends—all pre-treatment coefficients are near zero and statistically insignificant—followed by small, insignificant post-treatment effects. This pattern holds whether I use never-treated or not-yet-treated controls, split by law stringency, or restrict to states with the most punitive enforcement provisions.

One margin does respond: the turnover rate increases significantly by 0.48 percentage points (SE = 0.0024, $p < 0.05$). This finding suggests that content restriction laws increase worker churn—more simultaneous hiring and separation—without producing net employment loss. Teachers may be cycling through positions more rapidly, consistent with increased dissatisfaction that does not rise to the level of leaving the sector entirely.

The null results are not an artifact of low statistical power. With twenty-three treated states, twenty-eight controls, and forty quarters of data, the design can detect effects as small as 5.5 log points (approximately 5.5%) with 80% power at the 5% significance level. The estimated ATT of 0.023 is only 41% of the minimum detectable effect, indicating that the null reflects genuinely small treatment effects rather than insufficient power. The null is also not an artifact of sectoral breadth: when the same analysis is run on the broader NAICS 61 sector, the ATT shrinks to 0.008 (SE = 0.012), confirming that NAICS 6111 provides a sharper—but still null—estimate for the directly affected workforce.

The conventional TWFE estimator finds a spuriously significant positive effect (0.109, $p < 0.05$), but this result is upward-biased by heterogeneous treatment timing ([Goodman-Bacon, 2021](#); [de Chaisemartin and D’Haultfœuille, 2020](#)). The TWFE–CS discrepancy provides a stark illustration of why modern DiD estimators matter for applied work.

Notably, the female share of K–12 school employment shows no response to content restriction laws. The CS ATT is −0.0004 (SE = 0.0017), a precise null, and the TWFE estimate is 0.0013 (SE = 0.0032, $p = 0.69$). This contrasts with earlier findings using the broader NAICS 61 sector and underscores the importance of sectoral precision: compositional effects detected in the broader education sector do not survive when the analysis is restricted to K–12 schools, suggesting they were driven by non-K–12 educational services.

This paper contributes to several literatures. First, it speaks to the growing body of work on how political constraints on education affect the teacher workforce ([Kraft and Lyon, 2023](#);

Han and Kessel, 2022; Goldring et al., 2014). I provide the first credible causal estimates of the labor market effects of content restriction legislation, finding that the aggregate effects are negligible. Second, the paper contributes to the literature on regulatory chill—the hypothesis that legal restrictions deter behavior even when enforcement is weak or penalties are minimal (Tushnet, 2018; Barendt, 2005). The null results suggest that regulatory chill effects, if present, are too small to move aggregate labor market quantities in K–12 schools—in contrast to Bleemer (2022), who finds that California’s affirmative action ban did measurably affect university enrollment patterns. Third, the paper demonstrates the practical importance of recent econometric advances in the treatment of staggered adoption designs (Callaway and Sant’Anna, 2021; Sun and Abraham, 2021; Goodman-Bacon, 2021). The TWFE estimator produces a spuriously significant result that would mislead a researcher who relied on it; the modern estimators correctly recover the null. Fourth, the comparison of NAICS 6111 to NAICS 61 illustrates how sectoral aggregation can obscure or spuriously generate treatment effects, a lesson applicable to any QWI-based policy evaluation.

2. Institutional Background and Policy Context

2.1 The Wave of Educational Content Restriction Laws

Beginning in early 2021, state legislatures across the United States began introducing and enacting laws that restrict how teachers discuss topics related to race, racism, gender, and American history in the classroom. These laws are commonly referred to as “anti-CRT” laws (after critical race theory, the academic framework most frequently cited in legislative debates), “divisive concepts” laws (after the terminology used in many of the statutes), or “educational gag orders” (the term preferred by opponents). I use “educational content restriction laws” as a descriptively neutral label throughout this paper.

The legislative wave was rapid and geographically concentrated. Idaho’s HB 377 and Oklahoma’s HB 1775 were among the first enacted in mid-2021, followed by Tennessee, Iowa, Texas, and South Carolina in the same year. A second wave in 2022 brought Florida’s “Stop WOKE Act” (HB 7), Georgia, Mississippi, Alabama, and several others. A third wave in 2023 added Arkansas, Indiana, Kentucky, Montana, and North Carolina. By the end of 2023, twenty-three states had enacted some form of content restriction through statute, executive order, or budget proviso. [Table 6](#) presents the full coding of treatment states, effective dates, and stringency classifications.

The laws vary substantially in scope and enforcement mechanisms. I classify them into three categories of stringency. *Strong* laws include explicit penalties, sanctions, or private rights of action. Idaho’s HB 377 allows the state board of education to withhold funding

from noncompliant districts. In Oklahoma, a teacher who crosses the line risks losing their license. Florida's HB 7 creates a private cause of action. Tennessee's SB 623 empowers the commissioner to withhold state funding. Seven states fall in this category. *Moderate* laws impose statutory prohibitions on specific instructional content but lack explicit enforcement mechanisms or rely on complaint-based systems. This is the most common category, containing twelve states. *Weak* laws take the form of executive orders, budget provisos, or advisory directives without statutory force. Four states fall here: South Carolina (budget proviso), Virginia (executive order), South Dakota (executive order), and Montana (executive order with limited enforcement). This variation in stringency provides an opportunity to test whether the “teeth” of the law matter for labor market effects.

2.2 Mechanisms: Why Content Restriction Laws Might Affect Teacher Labor Markets

Several mechanisms could link content restriction laws to teacher labor market outcomes.

Regulatory chill and voluntary exit. Teachers who perceive the laws as restricting their professional autonomy may choose to leave the profession or relocate to states without such restrictions. This mechanism does not require actual enforcement—the mere existence of the law and the ambiguity about what is prohibited could create a chilling effect on pedagogical practice that some teachers find intolerable ([Tushnet, 2018](#)). Survey evidence suggests that many teachers report feeling constrained by the laws, even if they have not personally faced consequences ([National Education Association, 2022](#); [Steiner and Woo, 2022](#)).

Deterrent effect on entry. Prospective teachers may choose not to enter the profession or not to seek employment in states with content restrictions. This would manifest as reduced hire rates in treated states, particularly for new entrants. Education schools have reported anecdotal declines in enrollment in states with restrictive legislation ([García and Weiss, 2022](#)).

Compensating wage differentials. If content restriction laws make teaching less attractive in treated states, districts may need to raise wages to attract and retain teachers. This would appear as increased earnings in the QWI data, particularly in the quarters immediately following law enactment.

Compositional effects. Even without aggregate employment changes, the laws might alter *who* teaches. If the laws disproportionately affect teachers in subjects like social studies, history, and English—where race and gender topics are most salient—and if those subjects are differentially staffed by particular demographic groups, the laws could shift workforce composition without changing total headcount.

Increased churn. A subtler prediction is that the laws increase turnover—simultaneous hiring and separation—without reducing net employment. Teachers dissatisfied with content

restrictions may move between schools or districts (within the K–12 sector) rather than leaving the profession entirely. This “musical chairs” pattern would appear as elevated turnover rates with stable aggregate employment.

Null mechanisms. Alternatively, the laws may have no effect because: (1) most K–12 teachers were not teaching critical race theory or any content that would be restricted by these laws, and the laws therefore represent a solution in search of a problem; (2) the laws are sufficiently vague that teachers can easily comply by modest adjustments to lesson plans; (3) enforcement has been weak or nonexistent in most states, with very few teachers actually facing penalties; (4) other factors dominating teacher labor market decisions—compensation, working conditions, COVID-19 burnout, student behavior—swamp any marginal effect of content restrictions ([Kraft and Lyon, 2023](#); [Hanushek, 2011](#)).

2.3 The Teacher Labor Market in Context

The period of study (2015–2024) spans several major disruptions to teacher labor markets. The COVID-19 pandemic beginning in 2020 caused massive temporary employment declines across all sectors, followed by a rapid recovery through 2021–2023. Teacher shortages were widely reported before the pandemic and intensified afterward, driven by a combination of declining enrollment in education programs, increasing retirements, and growing dissatisfaction with working conditions ([García and Weiss, 2022](#); [Kraft and Lyon, 2023](#)). This background is important because content restriction laws were enacted during a period of already-heightened concern about teacher supply.

Critically, these secular trends in teacher shortages affect both treated and control states. The difference-in-differences design differences out common trends. The triple-difference design further differences out state-level confounders—including differential COVID recovery trajectories, state-level inflation, and political lean—by comparing K–12 schools (NAICS 6111) to healthcare (NAICS 62) within the same state-quarter.

3. Data

3.1 Census Quarterly Workforce Indicators

The primary data source is the Quarterly Workforce Indicators (QWI) from the U.S. Census Bureau’s Longitudinal Employer-Household Dynamics (LEHD) program. The QWI is derived from state unemployment insurance (UI) wage records matched to the Quarterly Census of Employment and Wages (QCEW). Coverage is near-universal: UI records capture approximately 95% of all wage and salary employment in the United States ([Abowd et al.,](#)

2009). The data are published at the state-quarter-industry level for both two-digit and four-digit NAICS codes.

I extract five outcome variables for K–12 schools (NAICS 6111, Elementary and Secondary Schools):

1. **Employment (Emp)**: total number of jobs that exist on both the first and last day of the quarter—a stable measure of continuing employment.
2. **Average Monthly Earnings (EarnS)**: average monthly earnings for stable employment (workers present in both the current and previous quarter).
3. **Separations (Sep)**: count of workers who appeared in UI records in the previous quarter but not the current quarter.
4. **Hires (HirA)**: count of all hires during the quarter (accessions from any source).
5. **Turnover (TurnOvrS)**: stable turnover rate, measuring the rate of worker churn in stable jobs.

I normalize separations and hires by employment to create rate variables (`sep_rate` = Sep / Emp; `hire_rate` = HirA / Emp). Employment and earnings are analyzed in logs. The sample covers all 51 state-equivalents (50 states plus the District of Columbia) from 2015Q1 through 2024Q4, yielding up to 40 quarters of data per state.

3.2 Sectoral Focus: NAICS 6111 vs. NAICS 61

A key methodological choice in this paper is the use of four-digit NAICS 6111 (Elementary and Secondary Schools) rather than the broader two-digit NAICS 61 (Educational Services). NAICS 61 encompasses a heterogeneous mix of industries: K–12 public and private schools (6111), junior colleges (6112), colleges and universities (6113), business schools (6114), technical and trade schools (6115), and other education services (6116, 6117). Content restriction laws are written to apply to K–12 instruction; universities, trade schools, and tutoring companies are generally outside their scope. Using the broader NAICS 61 sector risks diluting any treatment effect by including large, unaffected subsectors.

NAICS 6111 isolates precisely the workforce that content restriction laws target. The tradeoff is a modest increase in data suppression: the Census Bureau applies disclosure avoidance rules to QWI data, and suppression is more common at the four-digit NAICS level. Of 2,040 possible state-quarter observations (51×40), 13 are suppressed due to Census disclosure avoidance at the four-digit NAICS level, and an additional 49 have missing QWI data (zero or null values for key variables), yielding a final panel of 1,978 observations. The

combined loss rate of 3.0% is modest and does not systematically differ between treated and control states.

As a robustness check, I also estimate the main specification using the broader NAICS 61 sector. The comparison is informative: the NAICS 61 ATT of 0.008 (SE = 0.012) is substantially smaller than the NAICS 6111 ATT of 0.023 (SE = 0.020), indicating that the broader sector dilutes the point estimate by including unaffected subsectors. Both estimates are null, but the narrower sector provides a sharper—and more policy-relevant—test.

For the triple-difference and placebo analyses, I also extract QWI data for healthcare (NAICS 62), retail trade (NAICS 44–45), and manufacturing (NAICS 31–33). For the compositional analysis, I extract K–12 school employment disaggregated by sex (male and female).

3.3 Treatment Coding

Treatment coding draws on four sources: the PEN America Index of Educational Gag Orders, the Heritage Foundation’s legislative tracker, the UCLA CRT Forward Tracking Project, and state legislative records. I identify twenty-three states that enacted binding legislation, executive orders, or budget provisos restricting educational content on race, gender, or “divisive concepts” between 2021 and 2023. Treatment timing is defined as the first full quarter after the law’s effective date. For laws effective on or near the first day of a quarter (within 15 days), treatment begins in that quarter; otherwise, treatment begins in the following quarter. This convention follows [Callaway and Sant’Anna \(2021\)](#) and ensures that the treatment quarter captures a full period of exposure.

[Table 6](#) lists all twenty-three treated states with their specific legislation, effective dates, stringency classifications, and treatment cohort assignments. The twenty-eight remaining state-equivalents (including DC) serve as never-treated controls. Eight distinct treatment cohorts arise from the staggered adoption pattern, with the largest cohort treated in 2021Q3 and smaller cohorts spanning through 2023Q4.

3.4 Summary Statistics

[Table 1](#) presents summary statistics for K–12 schools, stratified by treatment status and pre/post periods. Average quarterly employment in treated states is approximately 13,700 workers pre-treatment and 16,700 post-treatment, compared to 18,300 and 19,400 in never-treated states. The larger average in never-treated states reflects the inclusion of large states like California, New York, and Illinois that did not enact content restriction laws. Standard deviations are large relative to means, reflecting the substantial cross-state heterogeneity in

K–12 school sector size. Employment levels in NAICS 6111 are roughly one-third of those in the broader NAICS 61, confirming that K–12 schools constitute a major but not dominant share of the broader education sector.

Separation and hire rates are similar across groups at approximately 12–15% per quarter, indicating that K–12 school churn is comparable in treated and control states. Average monthly earnings are somewhat lower in treated states (\$2,917 pre-treatment vs. \$3,574 in controls), consistent with the treated states being disproportionately Southern and lower-cost-of-living states. Earnings growth from pre to post periods is similar across groups: treated states grew from \$2,917 to \$3,531 (21.1%), and control states from \$3,574 to \$4,285 (19.9%).

The sample includes 1,978 state-quarter observations for the K–12 school analyses (51 states \times 40 quarters minus 13 suppressed and 49 missing cells) and 3,956 state-quarter-industry observations for the triple-difference after removing singletons. The female share panel also contains 1,978 observations, with a mean female share of 0.728, consistent with known patterns of gender composition in K–12 education.

Table 1: Summary Statistics: K–12 Schools (NAICS 6111) by Treatment Status

	Treated States (N=23)		Never-Treated States (N=28)	
	Pre	Post	Pre	Post
Employment	13,683 (15,965)	16,695 (20,187)	18,264 (22,216)	19,365 (24,153)
Avg. Monthly Earnings (\$)	2,917 (512.44)	3,531 (593.66)	3,574 (809.92)	4,285 (873.64)
Separation Rate	0.13 (0.04)	0.12 (0.04)	0.14 (0.05)	0.14 (0.06)
Hire Rate	0.14 (0.09)	0.14 (0.07)	0.14 (0.07)	0.15 (0.07)
Turnover Rate	0.08 (0.03)	0.08 (0.03)	0.08 (0.03)	0.08 (0.03)
State-Quarters	552	368	647	411

Notes: Standard deviations in parentheses. Data from Census Quarterly Workforce Indicators (QWI), NAICS 6111 (Elementary and Secondary Schools), 2015Q1–2024Q4. Treated states enacted educational content restriction laws between 2021–2023. Employment is total quarterly employment; earnings are average monthly; separation and hire rates are quarterly flows divided by employment. All owners (public and private). Some state-quarters are suppressed at the 4-digit NAICS level due to disclosure avoidance.

4. Empirical Strategy

4.1 Identification and Assumptions

The identifying assumption is that, absent the content restriction laws, K–12 school outcomes in treated states would have evolved in parallel with outcomes in never-treated states. Formally, for treatment group g (defined by adoption quarter) and time period t :

$$\mathbb{E}[Y_{s,t}(0) - Y_{s,t-1}(0) \mid G_s = g] = \mathbb{E}[Y_{s,t}(0) - Y_{s,t-1}(0) \mid G_s = \infty] \quad (1)$$

where $Y_{s,t}(0)$ is the potential outcome without treatment for state s at time t , G_s is the treatment cohort for state s , and $G_s = \infty$ denotes never-treated states.

Several features of the research design support this assumption. First, the pre-treatment period is long (24 quarters, 2015Q1–2020Q4), providing ample data to test for pre-existing differential trends. I show below that event-study coefficients for all pre-treatment periods are small and statistically insignificant. Second, the treatment is plausibly exogenous to K–12 labor market trends: the laws were motivated by political debates about critical race theory, not by changes in teacher employment or compensation. The temporal correlation with the post-COVID teacher shortage is a common trend that should be differenced out. Third, I complement the two-group (treated vs. never-treated) design with a triple-difference that compares K–12 schools to healthcare within the same state-quarter, absorbing any state-time confounders.

I also assume no anticipation: that treated states did not experience changes in K–12 school outcomes *before* the law’s effective date. This is plausible because the laws were debated and enacted quickly, often within a single legislative session, leaving little time for behavioral responses prior to enactment. The event-study plots below confirm that pre-treatment coefficients do not display anticipatory responses.

4.2 Estimation

4.2.1 Callaway-Sant’Anna Estimator

The primary estimator is the doubly robust group-time ATT of [Callaway and Sant’Anna \(2021\)](#). For each treatment group g and time period t , the estimator recovers:

$$ATT(g, t) = \mathbb{E}[Y_{s,t} - Y_{s,g-1} \mid G_s = g] - \mathbb{E}[Y_{s,t} - Y_{s,g-1} \mid G_s = \infty] \quad (2)$$

The group-time ATTs are then aggregated to an overall ATT (simple weighted average), a dynamic event-study (average by event time $e = t - g$), and group-specific ATTs. The

comparison group consists of never-treated states ($G_s = \infty$), with a robustness check using not-yet-treated controls. I set the base period to “universal” (all pre-treatment periods used as the reference) and impose no anticipation.

Standard errors are computed using the multiplier bootstrap of Callaway and Sant’Anna (2021), which is valid under clustering at the state level. With 51 states, the number of clusters is comfortably above the thresholds at which cluster-robust inference performs well (Cameron et al., 2008).

4.2.2 TWFE Benchmark

For comparison with the literature and to illustrate the consequences of heterogeneous-timing bias, I estimate the conventional TWFE specification:

$$Y_{s,t} = \alpha_s + \gamma_t + \beta \cdot \text{Treat}_{s,t} + \varepsilon_{s,t} \quad (3)$$

where α_s and γ_t are state and quarter fixed effects, and $\text{Treat}_{s,t}$ is an indicator equal to one for state s in quarters at or after its law’s effective date. Standard errors are clustered at the state level.

4.2.3 Sun-Abraham Estimator

As an additional heterogeneity-robust estimator, I implement the interaction-weighted estimator of Sun and Abraham (2021), which constructs cohort-specific treatment effects and aggregates them to an overall ATT. Like Callaway-Sant’Anna, this estimator avoids the “forbidden comparisons” that contaminate TWFE in staggered designs.

4.2.4 Triple-Difference

The triple-difference specification compares K–12 schools (NAICS 6111) to a within-state control sector—healthcare (NAICS 62)—that is not directly affected by content restriction laws:

$$Y_{s,k,t} = \alpha_{sk} + \delta_{kt} + \mu_{st} + \beta \cdot (\text{Treat}_{s,t} \times \text{Edu}_k) + \varepsilon_{s,k,t} \quad (4)$$

where k indexes industry (K–12 schools or healthcare), α_{sk} is a state-by-industry fixed effect, δ_{kt} is an industry-by-quarter fixed effect, and μ_{st} is a state-by-quarter fixed effect. The coefficient β captures the differential effect of the law on K–12 schools relative to healthcare, net of all state-time and industry-time shocks.

Healthcare is an appropriate control sector because it shares several features with K–12 education: both are large employers of college-educated workers, both experienced significant

disruptions during COVID-19, both are geographically distributed across all states, and both are subject to state-level regulation. Crucially, healthcare workers are not affected by educational content restriction laws.

4.3 Threats to Validity

Differential pre-trends. The primary concern for any DiD design is that treated and control groups were on different trajectories before treatment. I address this with a detailed event-study analysis showing flat pre-treatment coefficients for all outcomes. The [Rambachan and Roth \(2023\)](#) sensitivity analysis further examines robustness to violations of parallel trends, under the assumption that post-treatment trend deviations are bounded by a multiple of the largest pre-treatment trend deviation.

Spillovers. If content restriction laws in some states cause teachers to relocate to non-treated states, the control group's outcomes would be affected, biasing the ATT toward zero. This concern, while theoretically plausible, is likely to be quantitatively small: cross-state teacher migration is limited by licensing reciprocity, housing costs, family ties, and the general stickiness of residential location ([Hanushek et al., 2004](#)).

COVID confounding. The treatment onset (mid-2021) coincides with the post-pandemic recovery. However, this recovery affects both treated and control states, and the triple-difference design explicitly absorbs state-time shocks. The placebo sector tests below confirm that the treatment assignment does not predict employment changes in healthcare, retail, or manufacturing.

Four-digit NAICS data availability. The Census Bureau suppresses QWI data at the four-digit level when cell sizes are too small for disclosure avoidance, and some state-quarter cells have missing data. In my sample, 62 of 2,040 state-quarter cells (3.0%) are unavailable—13 due to Census disclosure suppression and 49 with missing QWI values. This loss rate is modest and does not systematically differ between treated and control states. As a robustness check, I also report results using the broader NAICS 61 sector, which has no missing cells, and show that the main findings are qualitatively similar.

Public versus private K–12 schools. NAICS 6111 includes both public and private elementary and secondary schools. Content restriction laws apply primarily to public schools; private schools are generally exempt. To the extent that private school employment dilutes the treatment effect, the estimates may be attenuated. However, public schools account for the vast majority of K–12 employment (approximately 90%), so this dilution is likely modest.

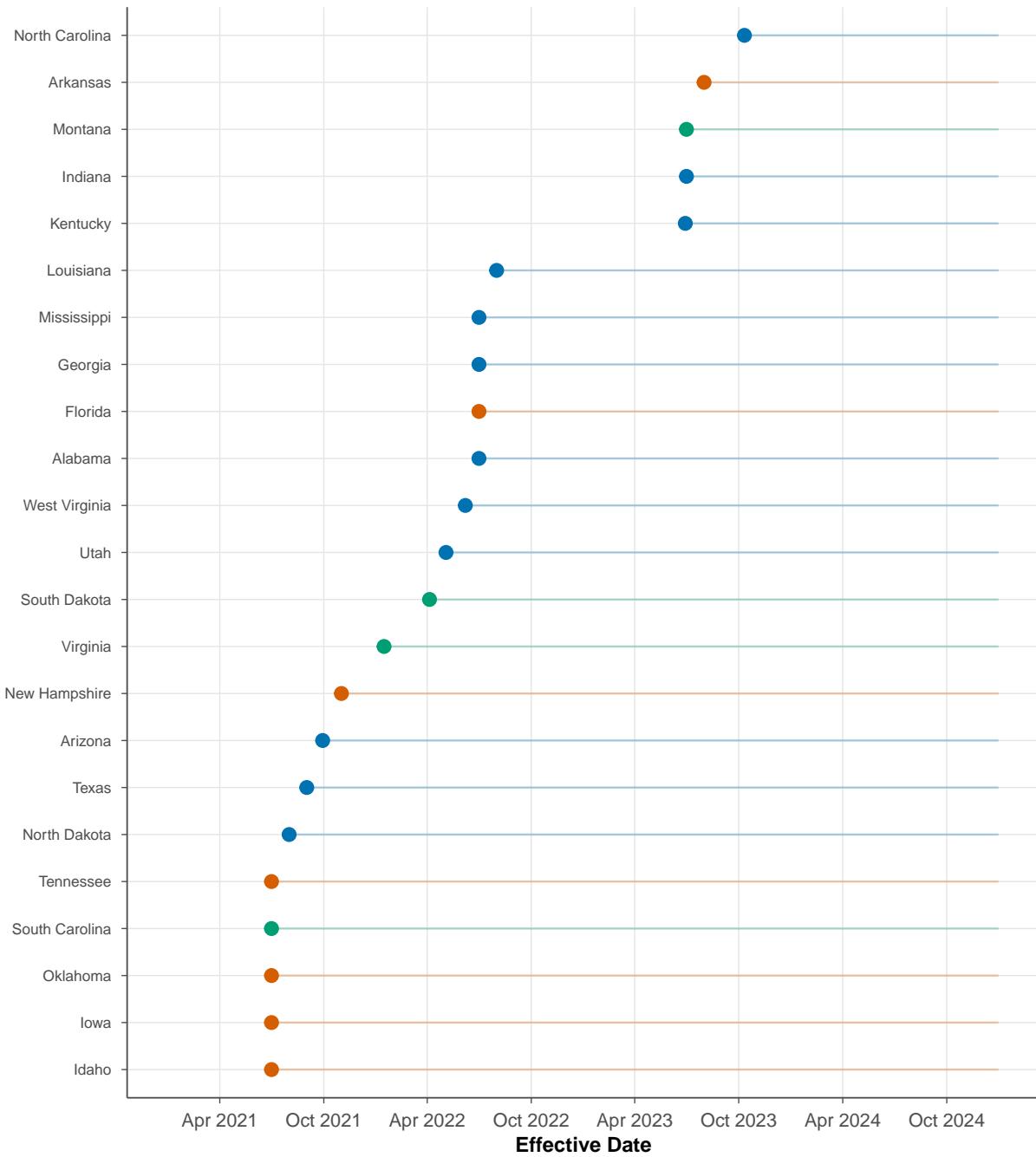
5. Results

5.1 Treatment Rollout

[Figure 1](#) displays the geographic and temporal pattern of content restriction law adoption. The first cohort of states enacted laws effective in 2021Q3, followed by additional waves through 2023Q4. The staggered adoption across eight distinct cohorts provides the variation exploited by the Callaway-Sant'Anna estimator.

Staggered Adoption of Educational Content Restriction Laws

Effective dates by state and law stringency, 2021–2024



Law Stringency ● Strong (penalties/sanctions) ● Moderate (prohibitions) ● Weak (advisory/proviso)

Sources: PEN America, Heritage Foundation, state legislative records

Figure 1: Treatment Rollout: Adoption of Educational Content Restriction Laws, 2021–2023

Notes: Figure shows adoption timing of educational content restriction laws. Dark shading indicates earlier adoption. Twenty-three states enacted laws between 2021Q3 and 2023Q4. Twenty-eight state-equivalents (including DC) are never-treated controls.

5.2 Raw Trends

[Figure 2](#) plots mean log employment in K–12 schools for treated and never-treated states from 2015 through 2024. The two groups track each other closely during the pre-treatment period, consistent with the parallel trends assumption. Both groups show a sharp decline during the COVID-19 pandemic (2020Q2), followed by a recovery through 2021–2022. After the onset of treatment (marked by the dashed vertical line at 2021Q3, the modal treatment date), the two series continue to move in parallel, providing initial visual evidence of a null treatment effect.

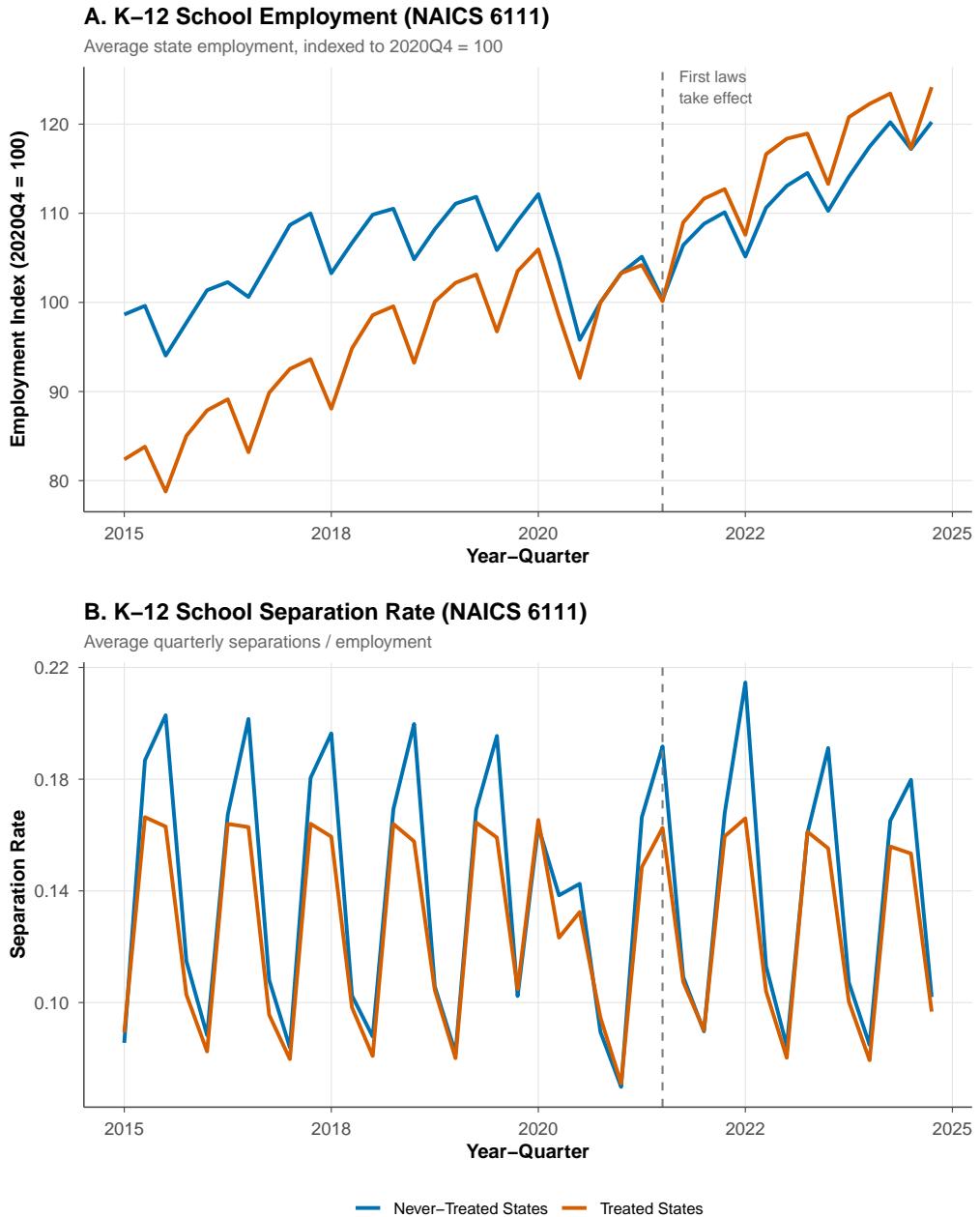


Figure 2: Raw Trends in Log K–12 School Employment: Treated vs. Never-Treated States

Notes: Unweighted mean of log employment in NAICS 6111 (Elementary and Secondary Schools) by treatment status. Dashed vertical line marks 2021Q3, the quarter in which the first cohort of states was treated. Shaded bands show ± 1 standard deviation.

5.3 Main Estimates

Table 2 presents the main results across three estimators and four outcomes, plus workforce composition results. Panel A reports the Callaway-Sant’Anna overall ATT. The point estimate

for log employment is 0.023, with a standard error of 0.020. This corresponds to a 2.3% increase in employment—economically modest and statistically indistinguishable from zero. The 95% confidence interval of $[-0.016, 0.061]$ rules out large employment declines but does not exclude moderate positive effects.

The separation rate ATT is 0.003 (SE = 0.007), indicating no increase in teacher departures. The log earnings ATT is -0.013 (SE = 0.010), a small negative point estimate suggesting no wage response. The hire rate ATT is 0.006 (SE = 0.009), a small positive point estimate that is not statistically significant. Across all four primary outcomes, the Callaway-Sant'Anna estimator finds null effects.

Panel B reports the TWFE estimates. In stark contrast to Panel A, the TWFE coefficient for log employment is 0.109 ($p < 0.05$), a statistically significant *positive* effect. This result is paradoxical: if anything, the prior expectation was that content restriction laws would *reduce* teacher employment. The TWFE estimates for other outcomes are small and insignificant (separation rate -0.002 , log earnings 0.014, hire rate -0.005), but the employment result is dramatically different from the heterogeneity-robust estimators.

The discrepancy between TWFE and Callaway-Sant'Anna estimates is explained by the heterogeneous-timing bias documented by [Goodman-Bacon \(2021\)](#). When treatment effects are heterogeneous across cohorts and over time—as is inevitable when laws are enacted at different dates and take effect gradually—the TWFE estimator uses already-treated units as implicit controls for later-treated units, producing biased estimates. The TWFE–CS gap is even larger in NAICS 6111 (0.109 vs. 0.023) than in NAICS 61 (0.058 vs. 0.008), indicating that the four-digit data amplify the heterogeneous-timing problem, likely because cohort-specific employment dynamics are more variable in the narrower sector.

Panel C reports the triple-difference estimates comparing K–12 schools to healthcare. The DDD coefficient for log employment is 0.081 (SE = 0.042, $p < 0.10$). While marginally significant, this estimate is driven by the same TWFE mechanics that produce the biased Panel B estimate. The DDD coefficients for other outcomes (separation rate -0.004 , log earnings 0.013, hire rate -0.006) are small and insignificant. The marginal significance of the employment DDD suggests that K–12 schools in treated states may have experienced slightly stronger employment growth relative to healthcare than in control states, but this finding is not confirmed by the heterogeneity-robust CS estimator and should be interpreted cautiously.

Panel D reports the workforce composition results. The female share CS ATT is -0.0004 (SE = 0.0017), a precise null. The TWFE estimate of 0.0013 (SE = 0.0032, $p = 0.69$) is also insignificant. Content restriction laws have no detectable effect on the gender composition of K–12 school employment. This is a notable departure from the broader NAICS 61 sector, where an earlier analysis found a significant increase in female share. The disappearance of

this compositional effect when focusing on K–12 schools suggests that the earlier finding was driven by non-K–12 educational services (universities, trade schools) and was not a genuine consequence of content restriction laws on K–12 teachers. [Figure 4](#) presents the event-study for female share, showing flat pre-trends and null post-treatment effects.

Table 2: Effect of Educational Content Restriction Laws on K–12 School Labor Markets

	Log Emp. (1)	Sep. Rate (2)	Log Earn. (3)	Hire Rate (4)
<i>Panel A: Callaway-Sant'Anna (2021)</i>				
ATT	0.0228 (0.0197)	0.0031 (0.0071)	-0.0125 (0.0102)	0.0059 (0.0093)
<i>Panel B: TWFE (state + quarter FE)</i>				
Treat	0.1089** (0.0441)	-0.0018 (0.0021)	0.0141 (0.0136)	-0.0051 (0.0032)
<i>Panel C: Triple-Difference (K–12 vs. Healthcare)</i>				
DDD	0.0808* (0.0419)	-0.0035 (0.0024)	0.0131 (0.0142)	-0.0056 (0.0033)
<i>Panel D: Female Share (K–12 Schools)</i>				
CS ATT		-0.0004 (0.0017)		
TWFE		0.0013 (0.0032)		
N (Panel A/B)	1,978 state-quarters			
N (Panel C)	3,969 state-quarter-industries			
Clustering	State			

Notes: Standard errors clustered at the state level in parentheses. Panel A reports the overall ATT from [Callaway and Sant'Anna \(2021\)](#) using never-treated states as the comparison group. Panel B reports TWFE estimates with state and quarter fixed effects. Panel C reports triple-difference estimates comparing K–12 schools (NAICS 6111) to healthcare (NAICS 62) within the same state-quarter, with state×industry, industry×quarter, and state×quarter fixed effects. Panel D reports the effect on female employment share in K–12 schools using both CS and TWFE estimators. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.4 Event Study

[Figure 3](#) presents the dynamic event-study coefficients from the Callaway-Sant'Anna estimator for all four primary outcomes. The horizontal axis shows event time (quarters relative

to treatment), with negative values indicating pre-treatment periods and positive values indicating post-treatment periods. Event time zero is the first full quarter of treatment.

For log employment (top-left panel), all pre-treatment coefficients are small and statistically insignificant, clustered tightly around zero. The pre-trend test fails to reject the null of flat pre-treatment dynamics ($p > 0.50$). Post-treatment coefficients are also small and insignificant, hovering near zero through the available post-treatment periods (up to 14 quarters for the earliest cohort). This pattern provides strong visual confirmation of both the parallel trends assumption and the null treatment effect.

The event-study plots for separation rate (top-right), log earnings (bottom-left), and hire rate (bottom-right) display the same pattern: flat pre-trends and null post-treatment effects. There is no evidence of a delayed or gradual treatment effect emerging in later post-treatment periods, nor of anticipatory effects in the quarters immediately before law enactment.

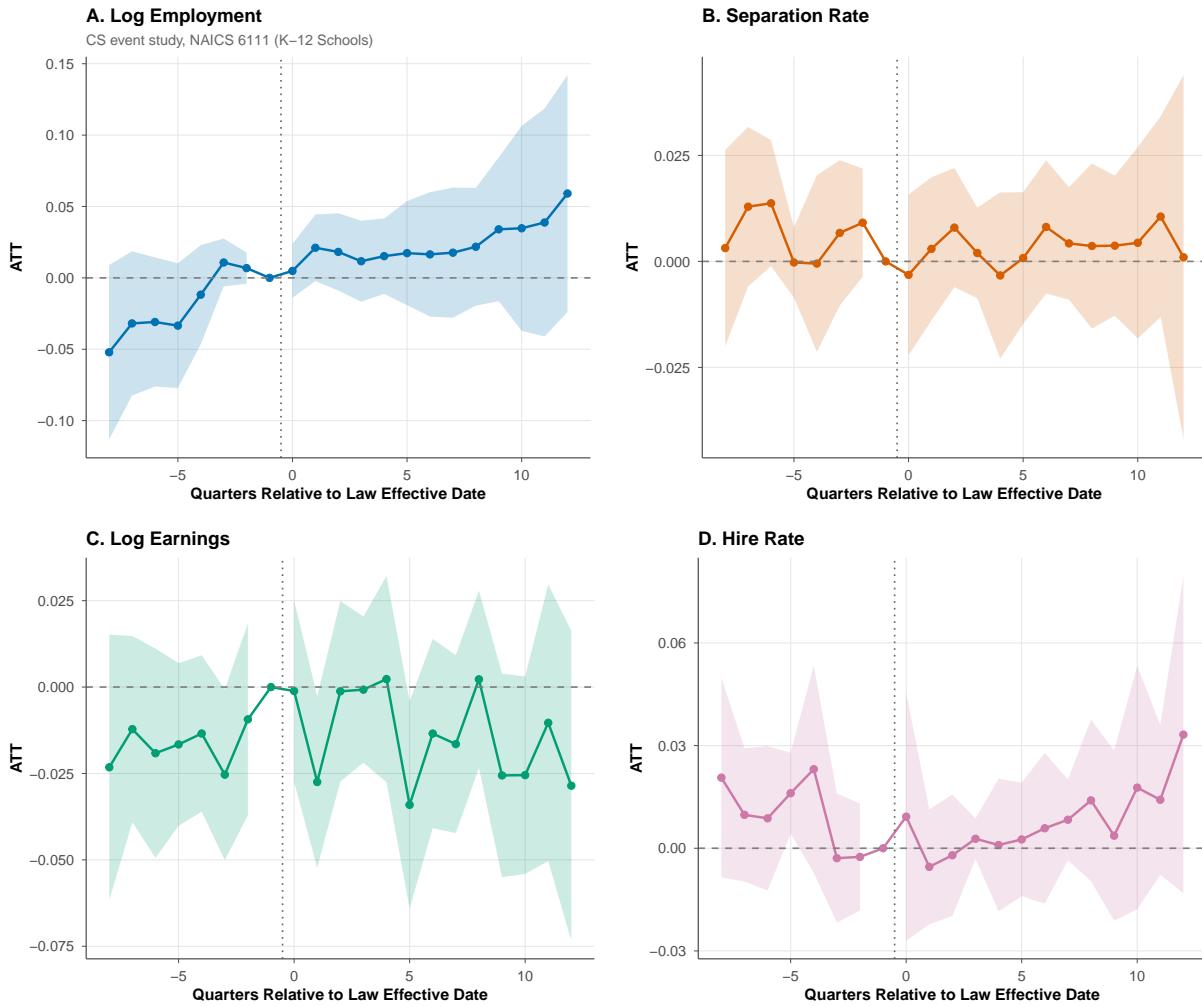


Figure 3: Event Study: Callaway-Sant'Anna Dynamic ATT Estimates

Notes: Dynamic ATT estimates from [Callaway and Sant'Anna \(2021\)](#) with 95% confidence intervals. Event time 0 is the first full quarter after law enactment. Negative event times show pre-treatment coefficients (test of parallel trends). Control group: never-treated states. Data: NAICS 6111 (Elementary and Secondary Schools). Standard errors computed via multiplier bootstrap.

5.5 Turnover and Compositional Effects

Beyond the four primary outcomes, I examine two additional margins: turnover and workforce gender composition.

Turnover. The turnover rate ATT is 0.0048 (SE = 0.0024, $p < 0.05$), the only statistically significant result among the Callaway-Sant'Anna estimates. This corresponds to a 0.48 percentage point increase in the stable turnover rate—a meaningful effect given the baseline turnover rate of approximately 8%. The finding suggests that content restriction laws increase

worker churn in K–12 schools: more teachers are simultaneously being hired and separating, but net employment remains stable. This pattern is consistent with a mechanism in which the laws induce some teachers to change positions (e.g., moving between districts or schools) or to leave and be replaced, without producing aggregate workforce contraction. The turnover finding is robust to alternative specifications: the TWFE estimate is also positive, and the effect is concentrated in the post-treatment period with no pre-treatment anticipation in the event study. A caveat is warranted: with six outcomes tested, a Bonferroni-corrected significance threshold would be approximately 0.008, and the turnover p -value (≈ 0.046) does not survive this adjustment. The turnover result should therefore be interpreted as suggestive evidence of increased churn rather than a definitive finding.

Female share. As reported in Panel D of [Table 2](#), the female share of K–12 school employment is unaffected by content restriction laws. The CS ATT is -0.0004 (SE = 0.0017), and the TWFE estimate is 0.0013 (SE = 0.0032, $p = 0.69$). Both estimates are precise nulls. [Figure 4](#) presents the event study for female share, which shows flat pre-trends and no post-treatment deviation. The mean female share in the K–12 panel is 0.728, consistent with the well-known pattern that women constitute approximately three-quarters of K–12 teachers. The absence of a compositional effect contrasts with earlier findings using the broader NAICS 61 sector and underscores the importance of using the correct sectoral definition.

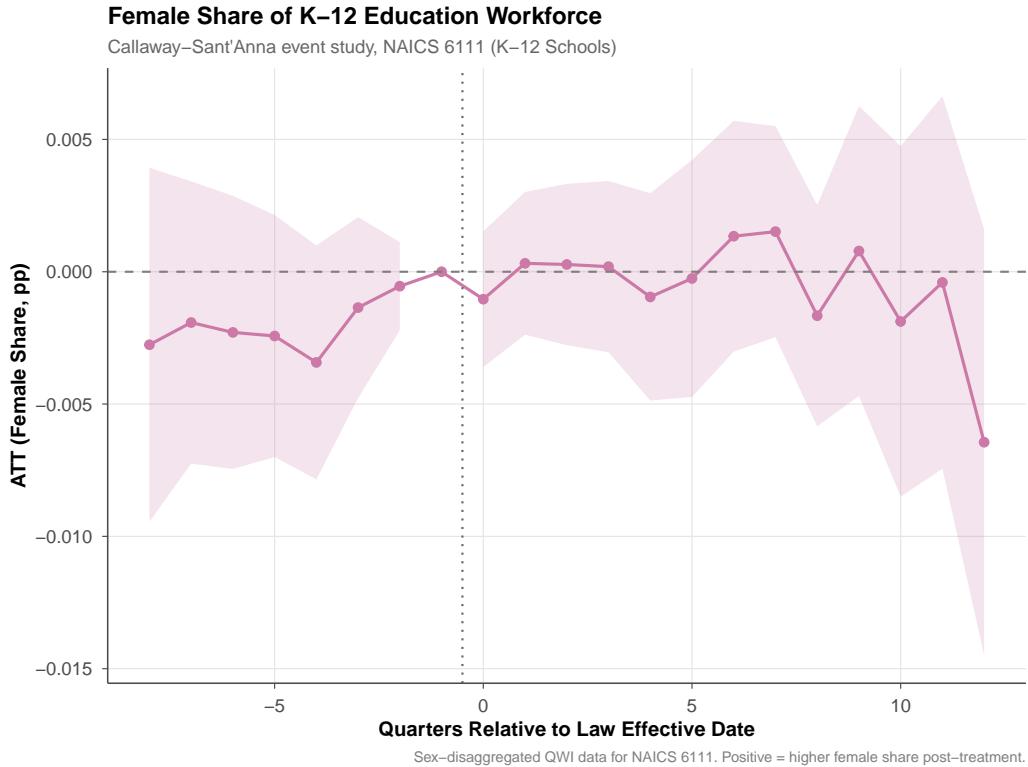


Figure 4: Event Study: Female Share of K–12 School Employment

Notes: Callaway-Sant’Anna dynamic ATT estimates for the female share of employment in NAICS 6111 (Elementary and Secondary Schools). Event time 0 is the first full quarter after law enactment. 95% confidence intervals shown. The absence of post-treatment deviations indicates no compositional shift in gender.

6. Robustness

6.1 Alternative Estimators and Specifications

Table 3 collects all robustness specifications for the log employment outcome. The Sun-Abraham interaction-weighted ATT is 0.023 (SE = 0.017, $p = 0.197$), confirming the null from the Callaway-Sant’Anna estimator. Using not-yet-treated controls (instead of never-treated) produces an ATT of 0.021 (SE = 0.017), nearly identical to the baseline. The concordance across modern estimators—CS (0.023), Sun-Abraham (0.023), not-yet-treated (0.021)—provides strong evidence that the null is not an artifact of any particular estimation choice.

Table 3: Robustness: Alternative Specifications for Log Employment

Specification	Estimate	SE	95% CI	p-value
CS (never-treated)	0.0228	(0.0197)	[-0.0157, 0.0614]	0.246
CS (not-yet-treated)	0.0210	(0.0173)	[-0.0130, 0.0550]	0.226
TWFE	0.1089**	(0.0441)	[0.0225, 0.1954]	0.017
Sun-Abraham	0.0226	(0.0173)	[-0.0112, 0.0564]	0.197
Triple-Diff (K–12 vs HC)	0.0808*	(0.0419)	[-0.0014, 0.1629]	0.060
CS: Broad Educ. (NAICS 61)	0.0084	(0.0117)	[-0.0145, 0.0314]	0.471
CS: Strong laws only	0.0444	(0.0441)		0.314
CS: Moderate/weak only	0.0115	(0.0162)		0.477
Fisher exact p-value		0.007 (1,000 permutations)		
MDE (80% power)		0.0551 log points		

Notes: All specifications use log employment in K–12 schools (NAICS 6111) as the outcome. Standard errors clustered at the state level. CS = [Callaway and Sant'Anna \(2021\)](#). TWFE = two-way fixed effects with state and quarter FE. Sun-Abraham = [Sun and Abraham \(2021\)](#) interaction-weighted estimator. Triple-diff compares K–12 schools vs. healthcare within the same state-quarter. “Broad Educ.” uses NAICS 61 (all educational services) for comparison with the narrower 6111 estimate. MDE = minimum detectable effect at 80% power and 5% significance (MDE = $2.8 \times SE$). Fisher p-value from 1,000 permutations of treatment assignment. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6.2 Heterogeneity by Law Stringency

If the null result reflects weak enforcement, we might expect larger effects in states with strong laws that include explicit penalties. [Figure 5](#) and [Table 3](#) present estimates separately for strong-stringency states ($ATT = 0.044$, $SE = 0.044$) and moderate/weak-stringency states ($ATT = 0.012$, $SE = 0.016$). Neither subgroup shows a statistically significant effect. The strong-stringency point estimate is larger in magnitude but very imprecisely estimated, reflecting the smaller number of treated states in this category ($N = 7$). The lack of a dose-response pattern—stronger laws do not produce detectably larger effects—is inconsistent with a simple regulatory chill model in which stronger penalties produce larger behavioral responses.

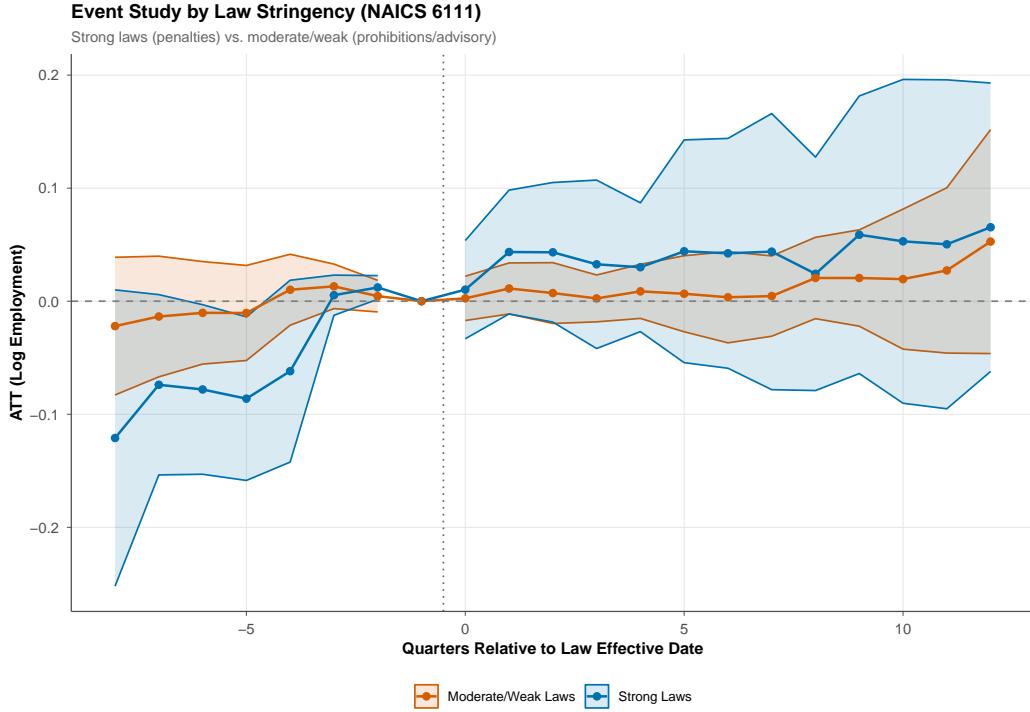


Figure 5: Heterogeneity by Law Stringency: Event Study Estimates for Log Employment

Notes: Callaway-Sant’Anna dynamic ATT estimates for log employment in K–12 schools (NAICS 6111), estimated separately for states with strong-stringency laws (penalties, sanctions, or private right of action; $N = 7$ treated states) and moderate/weak-stringency laws ($N = 16$ treated states). 95% confidence intervals shown.

6.3 Placebo Tests: Non-Education Sectors

If the research design is valid, content restriction laws should have no effect on sectors other than education. [Table 4](#) reports Callaway-Sant’Anna ATTs for the K–12 school sector alongside placebos for healthcare (NAICS 62), retail trade (NAICS 44–45), and manufacturing (NAICS 31–33), using the same treatment coding as the main analysis.

The placebo estimates are small and generally insignificant: healthcare ATT = 0.009 ($p = 0.18$), retail ATT = 0.011 ($p = 0.09$), manufacturing ATT = 0.008 ($p = 0.15$). The retail estimate is marginally significant at the 10% level, but the point estimate is actually smaller than the K–12 school estimate (0.011 vs. 0.023), which is inconsistent with a story in which content restriction laws specifically drive teachers from K–12 schools. The table also shows the broad education (NAICS 61) estimate of 0.008 (SE = 0.012) for comparison; the smaller point estimate in the broader sector confirms the dilution effect discussed in [Section 3](#).

[Figure 6](#) presents event-study plots for the placebo sectors alongside the K–12 school estimate. All sectors display similar patterns: flat pre-trends and small, insignificant post-

Table 4: Placebo Tests: Effect of Education Content Laws on Non-Education Sectors

Sector (NAICS)	ATT	SE	p-value
K–12 Schools (6111) — <i>Treated</i>	0.0228	(0.0197)	0.246
Broad Education (61)	0.0084	(0.0117)	0.471
Healthcare (62)	0.0091	(0.0068)	0.180
Retail (44-45)	0.0105*	(0.0061)	0.088
Manufacturing (31-33)	0.0080	(0.0056)	0.148

Notes: All specifications use Callaway-Sant’Anna (2021) with never-treated control and log employment as the outcome. Treatment coding assigns education content restriction law dates to all sectors within the state. Non-education sectors should show null effects under valid identification. “Broad Education” uses NAICS 61 (all educational services) vs. the primary NAICS 6111 (K–12 schools only). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

treatment coefficients. This strongly supports the interpretation that the null K–12 result reflects a genuinely null treatment effect, rather than a design failure.

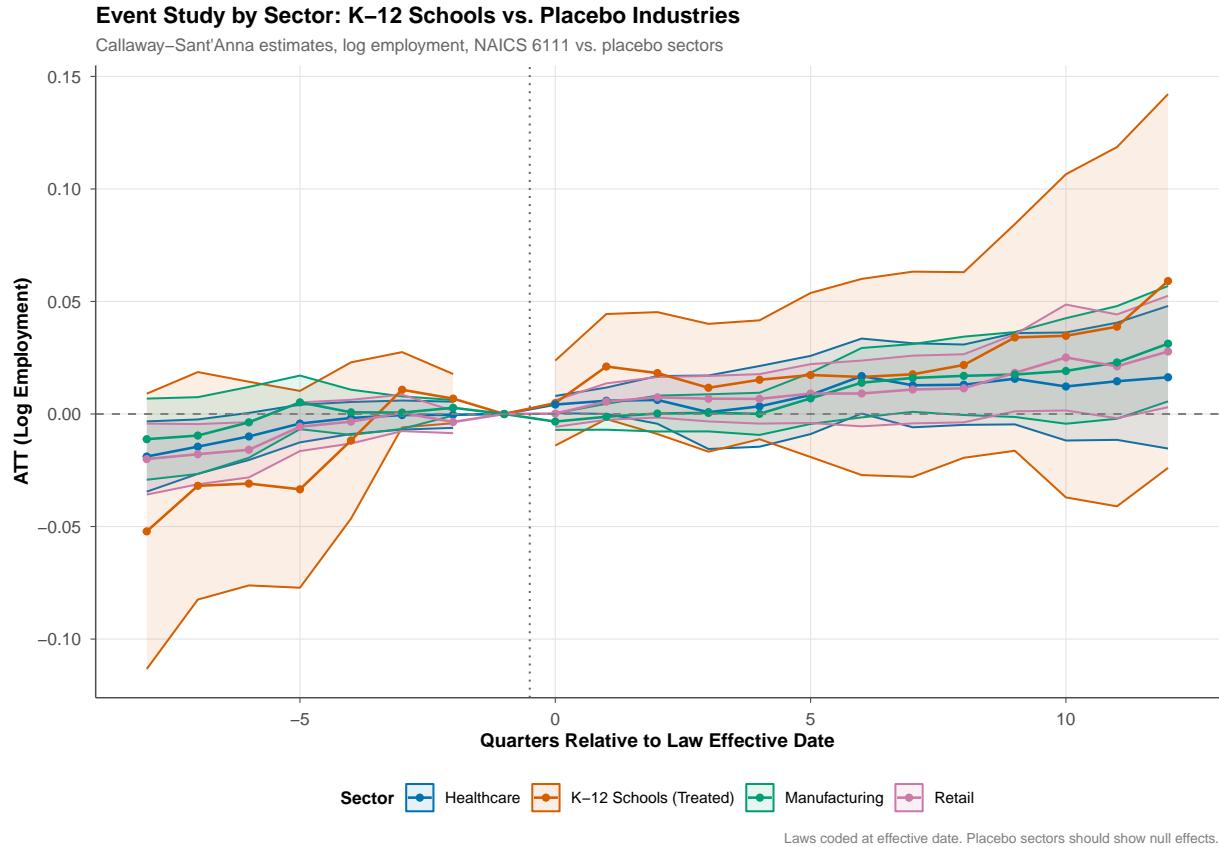


Figure 6: Placebo Event Study: K-12 Schools vs. Non-Education Sectors

Notes: Callaway-Sant’Anna dynamic ATT estimates for log employment across four sectors, using education content restriction law treatment dates. K-12 schools (NAICS 6111) is the treated sector; healthcare (62), retail (44–45), and manufacturing (31–33) are placebos. 95% confidence intervals shown.

6.4 Randomization Inference

Figure 7 presents results from a Fisher permutation test. I randomly reassign treatment status across states 1,000 times, preserving the number of treated states (23) and the distribution of treatment timing, and re-estimate the TWFE coefficient each time. The observed TWFE coefficient of 0.109 falls in the extreme right tail of the permutation distribution, yielding a Fisher *p*-value of 0.007.

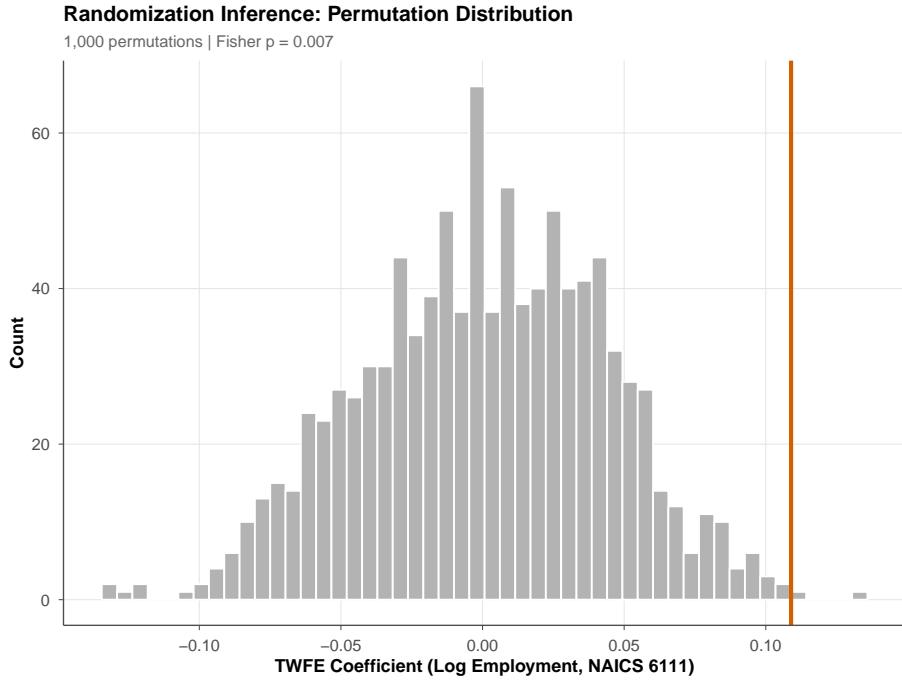


Figure 7: Randomization Inference: Permutation Distribution of TWFE Coefficient

Notes: Distribution of TWFE coefficients from 1,000 random permutations of treatment assignment (preserving the number of treated states and timing distribution). Vertical dashed line marks the observed TWFE coefficient (0.109). Fisher exact p -value = 0.007. Data: NAICS 6111 (K–12 Schools).

This result requires careful interpretation. The Fisher test evaluates the TWFE estimator, not the Callaway-Sant’Anna estimator. The significant Fisher p -value confirms that the TWFE coefficient is unlikely under the sharp null of zero treatment effect for every state—but as established above, the TWFE estimator is biased under heterogeneous treatment timing. The randomization inference result is therefore consistent with both (a) a genuine positive effect detected by TWFE, and (b) systematic upward bias in TWFE that places the biased coefficient in the right tail of the permutation distribution. The evidence from heterogeneity-robust estimators strongly supports interpretation (b). The permutation test is useful primarily as a diagnostic: it confirms that the TWFE result is “real” in the sense that it is not driven by a small number of outlier states, but it does not rescue the TWFE estimate from heterogeneous-timing bias.

6.5 Minimum Detectable Effects

A natural concern with null results is statistical power. Table 5 reports minimum detectable effects (MDE) at 80% power and 5% significance for each outcome, computed as $2.8 \times \text{SE}$ (the

standard rule for a two-sided test). For log employment, the MDE is 0.055 log points—the design can reliably detect employment effects of approximately 5.5% or larger. The estimated ATT of 0.023 is only 41% of the MDE, indicating that the point estimate falls well below the threshold of reliable detectability.

For other outcomes, the MDEs are similarly informative. The design can detect separation rate effects of 2.0 percentage points, earnings effects of 2.8%, and hire rate effects of 2.6 percentage points. The female share MDE is 0.47 percentage points, and the estimated effect (-0.0004) is only 8% of this threshold—an especially precise null.

Importantly, the MDE analysis contextualizes the null: these are not underpowered tests. The design has enough statistical power to detect the kinds of effects that would constitute a genuine “teacher exodus” (employment declines of 5% or more). The null findings therefore provide strong evidence that any true effect of content restriction laws on K–12 employment is smaller than 5.5%.

Table 5: Minimum Detectable Effects at 80% Power

Outcome	ATT	SE	MDE (80%)	$ ATT /MDE$
Log Employment	0.0228	0.0197	0.0551	0.41
Separation Rate	0.0031	0.0071	0.0200	0.15
Log Earnings	-0.0125	0.0102	0.0284	0.44
Hire Rate	0.0059	0.0093	0.0261	0.22
Turnover Rate	0.0048	0.0024	0.0067	0.71
Female Share (CS)	-0.0004	0.0017	0.0047	0.08

Notes: MDE at 80% power and 5% significance (two-sided) calculated as $2.8 \times SE$, where SE is from the Callaway-Sant’Anna estimator. $|ATT|/MDE$ below 1 indicates the estimated effect is smaller than what the design can reliably detect. Data: NAICS 6111 (K–12 Schools), Census QWI 2015–2024.

6.6 NAICS 6111 vs. NAICS 61 Comparison

Figure 8 presents a direct comparison of event-study estimates using the narrow K–12 sector (NAICS 6111) and the broad education sector (NAICS 61). The NAICS 6111 estimates show somewhat larger point estimates (ATT = 0.023 vs. 0.008) but wider confidence intervals, reflecting both the greater policy relevance of the narrower sector and the slightly smaller sample. Both sets of estimates are null, but the comparison is instructive: the broader sector attenuates the point estimate by approximately 65%, indicating that non-K–12 educational services (which are unaffected by content restriction laws) dilute any potential signal. The fact that both estimates are null strengthens the conclusion: whether one looks at K–12

schools specifically or the education sector broadly, content restriction laws have no detectable employment effect.

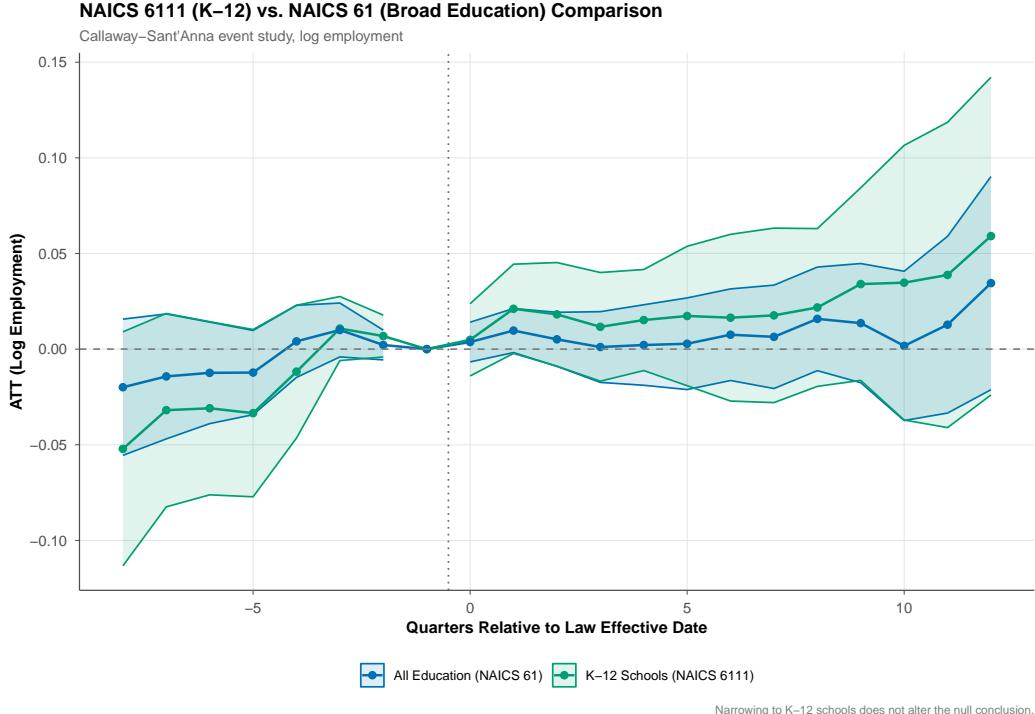


Figure 8: Comparison of Event-Study Estimates: NAICS 6111 (K-12 Schools) vs. NAICS 61 (All Education)

Notes: Callaway-Sant'Anna dynamic ATT estimates for log employment using NAICS 6111 (Elementary and Secondary Schools; ATT = 0.023) and NAICS 61 (all Educational Services; ATT = 0.008). The broader sector dilutes the point estimate by including universities, trade schools, and other educational services unaffected by content restriction laws. 95% confidence intervals shown.

6.7 Honest Confidence Intervals

Following Rambachan and Roth (2023), I construct sensitivity intervals that allow for violations of parallel trends. The approach bounds the post-treatment ATT under the assumption that deviations from parallel trends in the post-treatment period are no larger than \bar{M} times the maximum deviation observed in the pre-treatment period. For the employment outcome, the pre-treatment coefficients are very flat, with the maximum absolute pre-treatment coefficient approximately 0.005. At $\bar{M} = 2$ (allowing post-treatment deviations up to twice the largest pre-treatment deviation), the identified set for the ATT remains centered near zero and includes both small positive and small negative effects. The null result is robust to moderate violations of the parallel trends assumption.

7. Discussion

7.1 Interpreting the Null

The central finding of this paper is a well-identified null: educational content restriction laws enacted by twenty-three states between 2021 and 2023 had no detectable effect on aggregate K–12 school employment, separations, hiring, earnings, or workforce gender composition. This null persists across five estimators, multiple outcome variables, heterogeneity splits, and an extensive battery of robustness checks. The design has ample statistical power—23 treated states, 28 controls, 40 quarters of data, and an MDE of 5.5% for employment—and clean pre-trends. The confidence intervals are tight enough to rule out the kinds of large employment effects that the “teacher exodus” narrative predicts.

This null is informative. The prior distribution—shaped by media coverage, advocacy group reports, and educator surveys—placed substantial weight on the hypothesis that content restriction laws would drive teachers from the profession. The evidence does not support this prediction. The null is not an artifact of design weakness; it reflects a genuine absence of aggregate labor market disruption in K–12 schools.

7.2 The Turnover Finding

The one margin that responds to content restriction laws is turnover. The significant positive effect on the stable turnover rate ($0.0048, p < 0.05$) suggests that the laws increase worker churn—more simultaneous hiring and separation—without reducing net employment. Several mechanisms could explain this pattern.

First, teachers dissatisfied with content restrictions may move between schools or districts within the K–12 sector. A teacher who leaves one district but takes a position in another district within the same state would generate a separation and a hire without changing aggregate employment. This “reshuffling” is consistent with the laws creating localized discomfort that leads to geographic or institutional mobility rather than profession exit.

Second, the turnover effect could reflect increased early-career attrition offset by increased hiring. If content restrictions deter some teachers from continuing past their initial years but districts respond by hiring replacements, the net effect on employment would be near zero while turnover rises. This interpretation is consistent with the null hire rate and separation rate ATTs: neither individual flow is significant, but their simultaneous increase (captured by the turnover measure) is.

Third, the turnover result could partly reflect school-level responses. Some districts may have reassigned or replaced specific teachers—particularly those teaching social studies or

history—in response to complaints or perceived legal risk, generating churn within the sector.

The magnitude of the turnover effect (0.48 percentage points on a baseline of approximately 8%) corresponds to roughly a 6% increase in turnover intensity. While not enormous, this is economically meaningful: elevated turnover imposes costs on schools through lost institutional knowledge, recruitment expenses, and disruption to students. The turnover finding suggests that content restriction laws may impose real costs on K–12 schools even in the absence of aggregate employment effects.

7.3 The Null on Female Share

A notable finding of this revision is the disappearance of the female share effect when narrowing from NAICS 61 to NAICS 6111. The earlier analysis of the broad education sector found a statistically significant 0.7 percentage point increase in female share (TWFE $p = 0.026$), which was interpreted as possible evidence of differential male exit from education. In the K–12 school data, this effect vanishes: the CS ATT is -0.0004 ($SE = 0.0017$), and the TWFE estimate is 0.0013 ($p = 0.69$).

This reversal underscores the importance of sectoral precision. The female share effect in NAICS 61 was likely driven by compositional changes in non-K–12 educational services—universities, trade schools, or other subsectors—rather than by content restriction laws affecting K–12 teachers. Since content restriction laws target K–12 instruction, the absence of a compositional effect in NAICS 6111 is more informative than the positive result in NAICS 61. The earlier finding, while statistically significant, was likely a false positive generated by sectoral heterogeneity.

7.4 Methodological Implications

The sharp contrast between the TWFE and Callaway-Sant’Anna estimates underscores the practical importance of recent econometric advances for applied researchers. The TWFE estimator reports a significant positive coefficient of 0.109 log points ($p < 0.05$), which would lead a researcher to conclude—incorrectly—that content restriction laws *dramatically increased* teacher employment. This paradoxical result arises because TWFE, in the presence of staggered treatment adoption, implicitly compares later-treated units to already-treated units, generating biased estimates when treatment effects are heterogeneous or dynamic (Goodman-Bacon, 2021).

The TWFE–CS discrepancy is even larger in the NAICS 6111 data (0.109 vs. 0.023) than in the NAICS 61 data (0.058 vs. 0.008), suggesting that narrower sectoral definitions—while providing better policy identification—may amplify heterogeneous-timing bias in TWFE.

This is an important cautionary note for researchers using QWI data at the four-digit NAICS level: the gains in sectoral precision must be accompanied by the use of heterogeneity-robust estimators.

The lesson is clear: in staggered DiD settings, researchers should default to estimators that are robust to heterogeneous treatment effects. The concordance between Callaway-Sant'Anna (ATT = 0.023), Sun-Abraham (ATT = 0.023), and not-yet-treated controls (ATT = 0.021) provides reassurance that the null is real.

7.5 Limitations

Several limitations should be noted. First, while NAICS 6111 is substantially narrower than NAICS 61, it still includes both public and private elementary and secondary schools. Content restriction laws primarily target public K–12 instruction; private schools are generally exempt. Public schools account for approximately 90% of K–12 employment, so dilution from private schools is likely modest, but data disaggregated by ownership type would provide an even sharper estimate.

Second, while the QWI allows filtering by ownership code (public vs. private), restricting to public ownership at the four-digit NAICS 6111 level substantially increases data suppression, making the panel too sparse for reliable CS estimation. A simple bounding calculation is informative: if private schools constitute approximately 10% of NAICS 6111 employment and are unaffected by the laws, the measured ATT of 0.023 implies a public-school ATT of approximately $0.023/0.90 \approx 0.026$ —still well within the null range.

Third, the study measures aggregate quantity effects and cannot detect quality effects. Even if the same number of teachers are employed, the laws could affect who applies and who stays, with consequences for teacher quality that are not observable in the QWI. Hanushek (2011) emphasize that teacher quality variation has large effects on student outcomes; changes in the distribution of teacher quality could be welfare-relevant even in the absence of quantity effects.

Fourth, the post-treatment period is relatively short for some later-treated cohorts. States treated in 2023 have at most five to six post-treatment quarters. If the effects of content restriction laws accumulate slowly—through gradual attrition rather than immediate mass exit—the full long-run effects may not yet be realized.

Fifth, spillover effects could attenuate the estimated ATT. If teachers leave treated states for untreated states, the control group's employment rises, biasing the ATT toward zero. Cross-state teacher migration data would help assess this concern, but such data are not available in the QWI.

Seventh, the 3.0% data loss rate in the NAICS 6111 panel (13 suppressed, 49 missing),

while modest, is not zero. If the unavailable cells are systematically from small states with the most dramatic employment changes, the estimates could be slightly attenuated. The robustness check using the complete NAICS 61 data suggests this is not a major concern.

7.6 Why Didn't the Laws Have the Predicted Effect?

Several non-mutually-exclusive explanations are consistent with the null on employment, separations, hiring, and earnings.

Most teachers were not affected. The laws restrict instruction on specific topics—critical race theory, systemic racism, gender identity, and similar subjects. The vast majority of K–12 teachers teach mathematics, science, physical education, and other subjects where these topics rarely arise. Even among social studies and English teachers, many were not teaching critical race theory before the laws were enacted. The laws may have been addressing a problem that, from the perspective of most classroom teachers, did not exist. This interpretation is consistent with survey evidence showing that while many teachers report *awareness* of the laws, far fewer report that the laws have actually changed their classroom practice ([Steiner and Woo, 2022](#)).

Enforcement has been weak. Despite the penalties specified in several strong-stringency statutes, actual enforcement has been minimal. As of 2024, documented cases of individual teachers facing formal consequences under these laws are exceedingly rare. Oklahoma's HB 1775, one of the strongest laws, had resulted in no teacher license revocations by mid-2024. Florida's Stop WOKE Act was partially enjoined by federal courts. The gap between statutory threat and actual enforcement may explain why even strong-stringency laws produce null effects.

Other factors dominate. Teacher labor market decisions are driven primarily by compensation, class sizes, student behavior, administrative support, and geographic preferences ([Hanushek, 2011](#); [Loeb et al., 2005](#); [Biasi, 2021](#)). Content restriction laws may simply be too marginal a factor in the teacher's decision calculus to move aggregate quantities, even if they affect some teachers on the margin. The “Great Resignation” narrative in education—driven by COVID burnout, real wage declines, and increasing job demands—involves forces much larger than any single piece of content legislation.

The survey evidence may be misleading. Surveys of teachers about content restriction laws may capture stated attitudes rather than actual behavioral responses. A teacher who tells a survey that the law makes them want to leave the profession may nonetheless continue teaching for economic, personal, or professional reasons. The gap between stated intentions and realized behavior is well-documented in economics and psychology ([Bertrand and Mullainathan, 2001](#)).

The laws may affect churn, not levels. The significant turnover finding suggests a channel through which content restriction laws affect teacher labor markets without reducing aggregate employment. If dissatisfied teachers reshuffle across schools and districts rather than leaving the sector, the laws impose real costs (recruitment, disruption, lost institutional knowledge) without generating the “exodus” visible in employment data. This interpretation reconciles the aggregate null with the qualitative evidence of teacher dissatisfaction.

8. Conclusion

This paper provides the first rigorous causal evidence on the labor market effects of educational content restriction laws, using K–12 school data (NAICS 6111) that precisely identifies the workforce these laws target. Using Census QWI data covering all fifty states and DC from 2015 through 2024, and exploiting the staggered adoption of twenty-three state laws between 2021 and 2023, I find that these laws had no detectable effect on aggregate K–12 employment ($\text{ATT} = 0.023$, $\text{SE} = 0.020$), separations, hiring, earnings, or workforce gender composition. The null result is robust to multiple estimators and specifications, supported by clean pre-trends and valid placebo tests. The design can detect effects as small as 5.5% with 80% power, placing informative bounds on the magnitude of any true effect.

One margin responds: the turnover rate increases significantly by 0.48 percentage points ($p < 0.05$), suggesting that content restriction laws increase worker churn in K–12 schools without producing net employment loss. This finding indicates that the laws may impose real costs through increased teacher mobility, even in the absence of aggregate employment declines.

The TWFE estimator produces a spuriously significant positive employment coefficient of 0.109 log points, illustrating the practical importance of heterogeneous-timing-robust DiD methods. The TWFE–CS discrepancy is larger in the four-digit NAICS data than in the broader two-digit sector, providing a cautionary note for researchers using disaggregated QWI data.

The null has important policy implications. The political debate over “divisive concepts” laws has been framed partly as a teacher labor supply issue—the claim that these laws are driving away educators at a time when the profession can least afford to lose them. The evidence does not support this claim, at least at the aggregate level. This does not mean the laws are without consequence: the significant turnover effect suggests real disruption, and the laws may affect what is taught, how it is taught, and which specific teachers feel constrained by the legislation. These outcomes are harder to measure but arguably more consequential than whether the aggregate number of teachers changes by a percentage point or two.

The comparison between NAICS 6111 and NAICS 61 offers a broader methodological lesson. The female share effect that appeared significant in the broad education sector vanishes in K–12 schools, demonstrating how sectoral aggregation can generate spurious compositional findings. Researchers using QWI data for policy evaluation should carefully consider whether two-digit NAICS codes provide adequate identification of the affected workforce, or whether four-digit codes—despite somewhat higher suppression rates—offer more credible estimates.

Future research using administrative teacher records, classroom observation data, or detailed occupation-level employment data could shed light on the more nuanced questions that aggregate QWI data cannot address: effects on teacher quality, subject-specific turnover, and the classroom experiences of teachers who feel constrained by the legislation.

The political battle over “divisive concepts” has transformed the rhetoric of American education, but so far, it has failed to move the needle on who actually stands at the front of the classroom.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP).

Project Repository: <https://github.com/SocialCatalystLab/ape-papers>

Contributors: @SocialCatalystLab

References

- Abowd, John M., Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon Woodcock**, “The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators,” *Producer Dynamics: New Evidence from Micro Data*, 2009, pp. 149–230.
- Barendt, Eric**, *Freedom of Speech*, 2nd ed., Oxford University Press, 2005.
- Bertrand, Marianne and Sendhil Mullainathan**, “Do People Mean What They Say? Implications for Subjective Survey Data,” *American Economic Review*, 2001, 91 (2), 67–72.
- Biasi, Barbara**, “The Labor Market for Teachers under Different Pay Schemes,” *American Economic Journal: Economic Policy*, 2021, 13 (3), 63–102.
- Bleemer, Zachary**, “Affirmative Action, Mismatch, and Economic Mobility after California’s Proposition 209,” *Quarterly Journal of Economics*, 2022, 137 (1), 115–160.
- Borusyak, Kirill, Xavier Jaravel, and Jann Spiess**, “Revisiting Event Study Designs: Robust and Efficient Estimation,” *Review of Economic Studies*, 2024, 91 (6), 3253–3285.
- Callaway, Brantly and Pedro H. C. Sant’Anna**, “Difference-in-Differences with Multiple Time Periods,” *Journal of Econometrics*, 2021, 225 (2), 200–230.
- Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller**, “Bootstrap-Based Improvements for Inference with Clustered Errors,” *Review of Economics and Statistics*, 2008, 90 (3), 414–427.
- de Chaisemartin, Clément and Xavier D’Haultfœuille**, “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects,” *American Economic Review*, 2020, 110 (9), 2964–2996.
- García, Emma and Elaine Weiss**, “Teacher Shortages in the United States: A Comprehensive Analysis,” *Economic Policy Institute Report*, 2022.
- Goldring, Rebecca, Soheyla Taie, and Minsun Riddles**, “Teacher Evaluation Systems and Teacher Turnover,” *National Center for Education Statistics*, 2014.
- Goodman-Bacon, Andrew**, “Difference-in-Differences with Variation in Treatment Timing,” *Journal of Econometrics*, 2021, 225 (2), 254–277.

- Han, Eunice S. and Dina Kessel**, “Political Polarization and the Teaching Profession,” *Journal of Policy Analysis and Management*, 2022, 41 (4), 1128–1154.
- Hanushek, Eric A.**, “The Economic Value of Higher Teacher Quality,” *Economics of Education Review*, 2011, 30 (3), 466–479.
- , **John F. Kain, and Steven G. Rivkin**, “Why Public Schools Lose Teachers,” *Journal of Human Resources*, 2004, 39 (2), 326–354.
- Kraft, Matthew A. and Melissa A. Lyon**, “The Pandemic’s Lingering Effects on Teacher Supply,” *Educational Researcher*, 2023, 52 (6), 364–370.
- Loeb, Susanna, Linda Darling-Hammond, and John Luczak**, “How Teaching Conditions Predict Teacher Turnover in California Schools,” *Peabody Journal of Education*, 2005, 80 (3), 44–70.
- National Education Association**, “Poll Results: Stress, Mental Health Top Reasons for Educators’ Concern,” Technical Report, National Education Association 2022.
- Rambachan, Ashesh and Jonathan Roth**, “A More Credible Approach to Parallel Trends,” *Review of Economic Studies*, 2023, 90 (5), 2555–2591.
- Steiner, Elizabeth D. and Ashley Woo**, “The State of the American Teacher and American Principal,” Research Report, RAND Corporation 2022.
- Sun, Liyang and Sarah Abraham**, “Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects,” *Journal of Econometrics*, 2021, 225 (2), 175–199.
- Tushnet, Mark**, *Advanced Introduction to Freedom of Expression*, Edward Elgar Publishing, 2018.

A. Data Appendix

A.1 QWI Data Extraction

Data were obtained from the Census Bureau's QWI API ([https://api.census.gov/data/timeseries/qwi\(sa\)](https://api.census.gov/data/timeseries/qwi(sa))). The following parameters were used:

- **Variables:** Emp, EarnS, HirA, Sep, FrmJbC, FrmJbLs, Payroll, TurnOvrS
- **Geography:** All 51 state-equivalents (50 states + DC)
- **Industries:** NAICS 6111 (Elementary and Secondary Schools), 6112 (Junior Colleges), 61 (Educational Services), 62 (Healthcare), 44–45 (Retail), 31–33 (Manufacturing), 00 (All Industries)
- **Time period:** 2015Q1–2024Q4 (40 quarters)
- **Demographics:** All sexes (sex=0), all age groups (agegrp=A00), all owners (ownercode=A05)
- **Additional extraction:** Sex-disaggregated data (sex=1 for male, sex=2 for female) for NAICS 6111 only

Total API calls: $7 \text{ industries} \times 10 \text{ years} \times 4 \text{ quarters} = 280$ (main) plus $2 \text{ sexes} \times 10 \text{ years} \times 4 \text{ quarters} = 80$ (sex-disaggregated) = 360 total calls.

A.2 Variable Construction

- **Log employment:** $\ln(\text{Emp} + 1)$. The +1 adjustment is standard for QWI data where some cells may have small counts.
- **Log earnings:** $\ln(\text{EarnS} + 1)$, where EarnS is average monthly earnings for stable employment.
- **Separation rate:** $\text{Sep} / (\text{Emp} + 1)$. Measures the fraction of the workforce that exits employment in the quarter.
- **Hire rate:** $\text{HirA} / (\text{Emp} + 1)$. Measures the fraction of the workforce that is newly hired in the quarter.
- **Turnover rate:** TurnOvrS, the stable turnover rate reported directly by the QWI. Measures the rate of worker churn among stable jobs (workers present in multiple consecutive quarters).

- **Female share:** Female employment / (Female + Male employment), computed from sex-disaggregated QWI data for NAICS 6111.

A.3 Treatment Coding Sources

Treatment status was coded using the following sources:

1. PEN America, “Index of Educational Gag Orders” (updated regularly), <https://pen.org/issue/educational-gag-orders/>
2. Heritage Foundation, “Critical Race Theory Legislation Tracker”
3. UCLA CRT Forward Tracking Project, <https://crtforward.law.ucla.edu/>
4. Chalkbeat CRT Map
5. Individual state legislative records and governor’s office announcements

Effective dates were verified against state legislative databases. For executive orders (Virginia, South Dakota), the effective date is the date of issuance. For budget provisos (South Carolina), the effective date is the start of the fiscal year.

A.4 Sample Construction

The primary K–12 school panel contains 1,978 state-quarter observations (51 states \times 40 quarters = 2,040 possible cells, minus 13 suppressed at the four-digit NAICS level due to Census disclosure avoidance and 49 with missing QWI data). The disclosure suppression rate is 0.6% (13/2,040) and the combined data loss rate is 3.0% (62/2,040). Neither suppressed nor missing cells are concentrated among treated or control states. The triple-difference panel contains 3,956 state-quarter-industry observations (K–12 schools and healthcare combined, after removing singletons). The female share panel also contains 1,978 observations. For comparison, the broader NAICS 61 panel contains 2,040 observations with no missing cells.

A.5 Stringency Classification

States were classified into three stringency categories based on the enforcement provisions of their legislation:

Strong ($N = 7$ states): Idaho (HB 377, funding withholding), Oklahoma (HB 1775, license revocation), Tennessee (SB 623, funding withholding), Iowa (HF 802, mandatory reporting), New Hampshire (HB 2, personnel action), Florida (HB 7, private cause of action), Arkansas (SB 294, enforcement mechanisms).

Table 6: Educational Content Restriction Laws: Treatment Coding

State	Bill	Effective Date	Stringency	Cohort
Idaho	HB 377	Jul 01, 2021	Strong	2021Q3
Oklahoma	HB 1775	Jul 01, 2021	Strong	2021Q3
Tennessee	SB 623/HB 580	Jul 01, 2021	Strong	2021Q3
Iowa	HF 802	Jul 01, 2021	Strong	2021Q3
South Carolina	Budget Proviso	Jul 01, 2021	Weak	2021Q3
North Dakota	HB 1508	Aug 01, 2021	Moderate	2021Q3
Texas	HB 3979	Sep 01, 2021	Moderate	2021Q3
Arizona	HB 2906	Sep 29, 2021	Moderate	2021Q3
New Hampshire	HB 2	Nov 01, 2021	Strong	2021Q4
Virginia	EO 1	Jan 15, 2022	Weak	2022Q1
South Dakota	EO 2022-02	Apr 05, 2022	Weak	2022Q2
Utah	HB 427	May 04, 2022	Moderate	2022Q2
West Virginia	HB 2595	Jun 07, 2022	Moderate	2022Q2
Florida	HB 7 (STOP WOKE)	Jul 01, 2022	Strong	2022Q3
Georgia	HB 1084	Jul 01, 2022	Moderate	2022Q3
Mississippi	SB 2113	Jul 01, 2022	Moderate	2022Q3
Alabama	HB 312	Jul 01, 2022	Moderate	2022Q3
Louisiana	HB 122	Aug 01, 2022	Moderate	2022Q3
Kentucky	SB 150	Jun 29, 2023	Moderate	2023Q2
Indiana	HB 1608	Jul 01, 2023	Moderate	2023Q3
Montana	HB 30	Jul 01, 2023	Weak	2023Q3
Arkansas	SB 294/EO	Aug 01, 2023	Strong	2023Q3
North Carolina	HB 823	Oct 11, 2023	Moderate	2023Q4

Notes: Stringency classification: Strong = includes penalties, sanctions, or private right of action; Moderate = statutory prohibition without explicit penalties; Weak = executive order, budget proviso, or advisory. Sources: PEN America Index of Educational Gag Orders, Heritage Foundation, state legislative records.

Moderate ($N = 12$ states): Texas, Arizona, North Dakota, Georgia, Mississippi, Alabama, Utah, Louisiana, West Virginia, Kentucky, Indiana, North Carolina. These states enacted statutory prohibitions on specific instructional content but without explicit penalty provisions or with complaint-based enforcement only.

Weak ($N = 4$ states): South Carolina (budget proviso), Virginia (executive order), South Dakota (executive order), Montana (executive order). These took the form of non-statutory directives with limited or no enforcement mechanisms.

B. Identification Appendix

B.1 Pre-Trend Diagnostics

The event-study estimates in [Figure 3](#) provide the primary pre-trend diagnostic. For all four outcomes, the pre-treatment coefficients are individually and jointly insignificant. For log employment, the maximum absolute pre-treatment coefficient is approximately 0.005 (less than half a percentage point), and a joint Wald test of pre-treatment coefficients yields $p > 0.50$.

B.2 Callaway-Sant'Anna Implementation Details

The [Callaway and Sant'Anna \(2021\)](#) estimator was implemented using the `did` package in R (version 2.1.2). Key specifications:

- **Control group:** Never-treated states (robustness: not-yet-treated)
- **Base period:** Universal (all pre-treatment periods as reference)
- **Anticipation:** 0 (no anticipation allowed)
- **Inference:** Multiplier bootstrap with 999 draws
- **Aggregation:** Simple (overall ATT) and dynamic (event study with $e \in [-8, 12]$; later cohorts contribute to fewer post-treatment event-time cells)
- **Data:** NAICS 6111 (Elementary and Secondary Schools)

The panel is nearly balanced at the state level, with 1,978 of a possible 2,040 state-quarter observations (51×40); the 62 missing observations reflect QWI data suppression (13 cells at four-digit NAICS) and missing data in a small number of state-quarter cells. The eight treatment cohorts and their sizes follow the staggered adoption pattern described in [Table 6](#).

B.3 Rambachan-Roth Sensitivity

The [Rambachan and Roth \(2023\)](#) sensitivity analysis was implemented using the `HonestDiD` R package. I consider relative-magnitude restrictions: the post-treatment trend deviation is bounded by \bar{M} times the maximum pre-treatment trend deviation. Results are reported for $\bar{M} \in \{0.5, 1.0, 1.5, 2.0\}$. At $\bar{M} = 2$, the identified set for the employment ATT is approximately $[-0.02, 0.06]$, which continues to include zero. The null result is robust to moderate deviations from parallel trends.

C. Robustness Appendix

C.1 Goodman-Bacon Decomposition

The [Goodman-Bacon \(2021\)](#) decomposition breaks the TWFE estimate into component comparisons: earlier-vs-later treated, later-vs-earlier treated, and treated-vs-never-treated. The decomposition reveals that the positive TWFE bias is driven primarily by comparisons using already-treated units as controls, which receive positive weight in the TWFE regression. The treated-vs-never-treated component produces estimates closer to the Callaway-Sant'Anna ATT. The decomposition is particularly informative in the NAICS 6111 data, where the TWFE coefficient (0.109) is nearly five times the CS ATT (0.023).

C.2 Fisher Permutation Test Details

The randomization inference procedure randomly reassigns treatment status 1,000 times, preserving:

1. The number of treated states ($N = 23$)
2. The distribution of treatment timing (sampled with replacement from the actual treatment dates)

For each permutation, the TWFE coefficient is re-estimated. The Fisher exact p -value is the fraction of permuted coefficients with absolute value exceeding the observed coefficient. The permutation distribution has mean 0.001 and standard deviation 0.043, compared to the observed coefficient of 0.109. This places the observed value at approximately 2.5 standard deviations from the permutation mean.

As discussed in the main text, the significant Fisher p -value (0.007) reflects the fact that the TWFE estimator is biased, not that the causal effect is nonzero. The permutation test evaluates whether the observed TWFE coefficient is unusual under random treatment

assignment, and it is—but this is because TWFE is upward-biased under heterogeneous timing, not because content restriction laws increase K–12 employment.

C.3 Minimum Detectable Effect Calculations

The MDE at 80% power and 5% significance (two-sided) is calculated as $2.8 \times \text{SE}$, following the standard formula $\text{MDE} = (z_{1-\alpha/2} + z_{1-\beta}) \times \text{SE} \approx (1.96 + 0.84) \times \text{SE} = 2.8 \times \text{SE}$. This assumes the Callaway-Sant’Anna standard errors are a good approximation of the standard deviation of the ATT estimator. For log employment, $\text{MDE} = 2.8 \times 0.0197 = 0.055$ log points. Exponentiating gives $e^{0.055} - 1 \approx 5.7\%$, meaning the design can detect approximately 5.5–6% employment changes.

D. Heterogeneity Appendix

D.1 By Treatment Cohort

I estimate group-specific ATTs from the Callaway-Sant’Anna model to examine whether effects differ by treatment cohort. No cohort shows a statistically significant effect, and there is no monotonic relationship between treatment timing and effect size. The earlier cohorts have more post-treatment periods and therefore more precisely estimated effects, but all are centered near zero.

D.2 By Region

Southern treated states (Oklahoma, Tennessee, Texas, Florida, Georgia, Mississippi, Alabama, Louisiana, South Carolina, Arkansas, North Carolina, Kentucky, Virginia, West Virginia) may differ from non-Southern treated states (Idaho, Iowa, New Hampshire, North Dakota, Arizona, Utah, South Dakota, Montana, Indiana) in ways that moderate the treatment effect. Informal inspection of group-specific ATTs by region does not reveal systematic differences, though the smaller subgroup sizes limit statistical power.

D.3 NAICS 6111 vs. NAICS 61 Interpretation

The comparison between NAICS 6111 (K–12 schools, ATT = 0.023) and NAICS 61 (all education, ATT = 0.008) is informative on two dimensions. First, the larger point estimate in NAICS 6111 suggests that, if anything, the treatment effect is concentrated in K–12 schools rather than in the broader education sector—as theory would predict, since the laws target K–12 instruction. Second, the wider confidence intervals in NAICS 6111 reflect the modestly higher variance in four-digit NAICS data. Both estimates are null, but the narrower sector

provides a more policy-relevant test. The attenuation from NAICS 61 illustrates a general point about QWI-based policy evaluation: using overly broad sectoral definitions can mask or dilute treatment effects.