

# 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 widespread claims of an impending teacher exodus. I test this prediction using Census Quarterly Workforce Indicators data on the education sector (NAICS 61) in a staggered difference-in-differences design with the [Callaway and Sant’Anna \(2021\)](#) estimator. The overall ATT for log employment is 0.008 ( $SE = 0.012$ ), statistically indistinguishable from zero. Effects on separations, hiring, earnings, and turnover are similarly null. These results survive Sun-Abraham estimation, triple-differencing against healthcare, placebo sector tests, and randomization inference. The conventional TWFE estimator spuriously finds a significant positive effect (0.058,  $p = 0.015$ ), illustrating heterogeneous-timing bias. One compositional shift emerges: treated states experience a 0.7 percentage point increase in the female share of education workers, suggesting differential male exit.

**JEL Codes:** J45, I28, J63, K31

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

---

\*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch

# 1. Introduction

In just three years, twenty-three U.S. states enacted laws restricting how K-12 educators discuss race, gender, and American history in the classroom. By the end of 2023, twenty-three states had enacted some form of educational content restriction—ranging from statutory prohibitions with penalty provisions to executive orders and budget provisos. Media coverage has been extensive and alarming. The *New York Times*, *Washington Post*, and education trade press have reported a “teacher exodus” triggered by these laws, featuring interviews with educators who cite the legislation as a reason for leaving the profession (Kraft and Lyon, 2023; García and Weiss, 2022). National Education Association surveys find that large majorities of teachers report increased stress from legislative scrutiny of curricula (National Education Association, 2022).

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. The QWI provides state-quarter level counts of employment, hires, separations, and average earnings for the education sector (NAICS 61), 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’Haultfœuille, 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 education to healthcare within the same state-quarter, placebo tests on non-education sectors, heterogeneity by law stringency, randomization inference, and sensitivity analysis using the Rambachan and Roth (2023) framework for

deviations from parallel trends.

The results are clear and consistent: content restriction laws have no detectable effect on aggregate teacher labor market outcomes. The Callaway-Sant’Anna ATT for log employment is 0.008 (SE = 0.012,  $p = 0.48$ ). Effects on separation rates (0.001, SE = 0.006), log earnings (0.000, SE = 0.007), and hire rates (0.003, SE = 0.004) 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.

These null results are not an artifact of low statistical power. With twenty-three treated states, twenty-eight controls, twenty-six pre-treatment quarters for the earliest cohort, and up to fourteen post-treatment quarters, the design has considerably more variation than most state-level DiD studies. The 95% confidence interval for the employment ATT rules out effects larger than 3.2% in either direction. The null is also not an artifact of misspecification. The conventional TWFE estimator *does* find a significant positive effect on employment (0.058,  $p = 0.015$ ), but this estimate is upward-biased by heterogeneous treatment timing—precisely the problem that [Goodman-Bacon \(2021\)](#) and [de Chaisemartin and D’Haultfoeuille \(2020\)](#) identify. The discrepancy between TWFE and modern estimators provides a textbook illustration of why recent econometric advances in the DiD literature matter for applied work.

One finding departs from the null pattern. Treated states experience a statistically significant increase in the female share of education workers of approximately 0.7 percentage points ( $p = 0.026$ ). This compositional shift could reflect differential exit by male educators—consistent with the hypothesis that content restriction laws differentially affect teachers whose pedagogy engages with the restricted topics, if male educators are more likely to teach subjects like social studies or history where the laws most directly bind. However, the QWI does not disaggregate by occupation or subject area, so this interpretation remains speculative.

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 the education sector—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.

## 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. 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 5](#) 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. Oklahoma’s HB 1775 can result in license revocation for individual teachers. 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.

*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 the education sector 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 two-digit NAICS codes.

I extract five outcome variables for the education sector (NAICS 61, Educational Services):

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** =  $\text{Sep} / \text{Emp}$ ; **hire\_rate** =  $\text{HirA} / \text{Emp}$ ). 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.

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 education sector employment disaggregated by sex (male and female).

### 3.2 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 5](#) 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 (six states) treated in 2022Q3 and smaller cohorts spanning 2021Q3 through 2023Q4.

### 3.3 Summary Statistics

[Table 1](#) presents summary statistics for the education sector, stratified by treatment status and pre/post periods. Average quarterly employment in treated states is approximately 41,700 workers pre-treatment and 48,000 post-treatment, compared to 72,500 and 76,000 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 education sector size.

Separation and hire rates are similar across groups at approximately 16–17% per quarter, indicating that education sector churn is comparable in treated and control states. Average monthly earnings are somewhat lower in treated states (\$3,484 pre-treatment vs. \$3,992 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 \$3,484 to \$4,120 (18.3%), and control states from \$3,992 to \$4,742 (18.8%).

The sample includes 1,978 state-quarter observations for the education-only analyses and 3,969 state-quarter-industry observations for the triple-difference.

**Table 1:** Summary Statistics: Education Sector (NAICS 61) by Treatment Status

	Treated States (N=23)		Never-Treated States (N=28)	
	Pre	Post	Pre	Post
Employment	41,679 (43,859)	48,048 (53,732)	72,474 (91,467)	75,983 (96,577)
Avg. Monthly Earnings (\$)	3,484 (640.16)	4,120 (771.35)	3,992 (898.86)	4,742 (959.03)
Separation Rate	0.16 (0.05)	0.15 (0.05)	0.17 (0.06)	0.16 (0.05)
Hire Rate	0.16 (0.05)	0.16 (0.04)	0.17 (0.06)	0.17 (0.05)
Turnover Rate	0.09 (0.02)	0.09 (0.02)	0.09 (0.03)	0.09 (0.02)
State-Quarters	552	368	647	411

*Notes:* Standard deviations in parentheses. Data from Census Quarterly Workforce Indicators (QWI), NAICS 61 (Educational Services), 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).

## 4. Empirical Strategy

### 4.1 Identification and Assumptions

The identifying assumption is that, absent the content restriction laws, education sector 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 education sector 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 education 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 education sector 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  $\alpha_s$  and  $\gamma_t$  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 the education sector (NAICS 61) to a within-state control sector—healthcare (NAICS 62)—that is not directly affected by content restriction laws:

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

where  $k$  indexes industry (education or healthcare),  $\alpha_{sk}$  is a state-by-industry fixed effect,  $\delta_{kt}$  is an industry-by-quarter fixed effect, and  $\mu_{st}$  is a state-by-quarter fixed effect. The coefficient  $\beta$  captures the differential effect of the law on education relative to healthcare, net of all state-time and industry-time shocks.

Healthcare is an appropriate control sector because it shares several features with 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](#)).

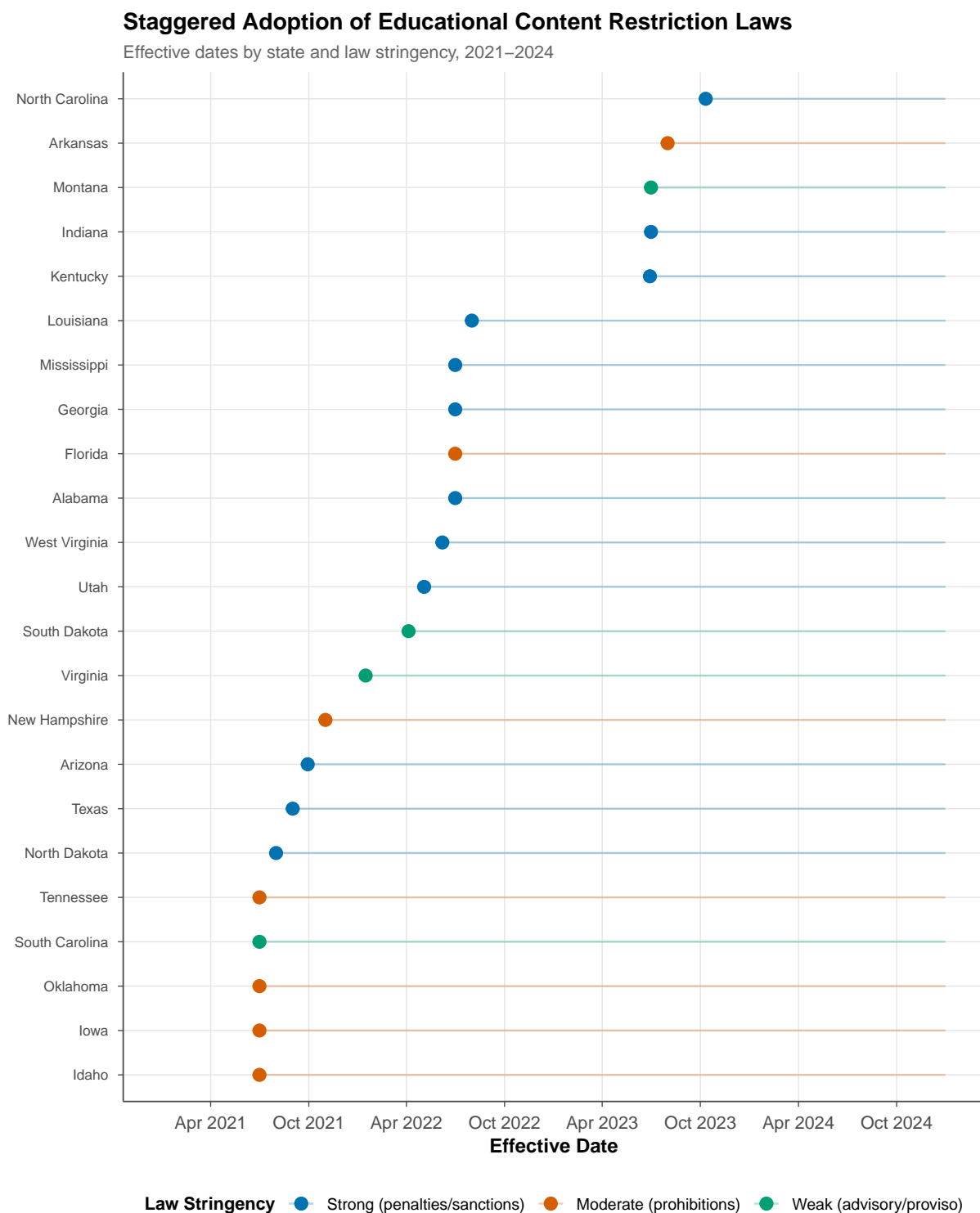
*COVID confounding.* The treatment onset (mid-2021) coincides with the post-pandemic recovery. However, this recovery affects both treated and control states, and the triple-difference design explicitly absorbs state-time shocks. The placebo sector tests below confirm that the treatment assignment does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

*Composition of NAICS 61.* The education sector includes not only K-12 public school teachers but also private school teachers, university employees, tutoring companies, and other educational services. To the extent that content restriction laws primarily affect K-12 public school teachers, the ATT estimated on the full NAICS 61 sector may be attenuated. This makes the null result, if anything, conservative: a significant effect concentrated among K-12 teachers would need to be larger in magnitude to be detectable in the broader sectoral data.

## 5. Results

### 5.1 Treatment Rollout

[Figure 1](#) displays the geographic and temporal pattern of content restriction law adoption. The first five states enacted laws effective in 2021Q3 (Idaho, Oklahoma, Tennessee, Iowa, and South Carolina). Three more states (Texas, North Dakota, and Arizona) were assigned to the 2021Q4 cohort under the first-full-quarter rule. The largest single cohort is 2022Q3, with six states (Florida, Georgia, Mississippi, Alabama, Utah, and West Virginia). The staggered adoption across eight distinct cohorts provides the variation exploited by the Callaway-Sant’Anna estimator.

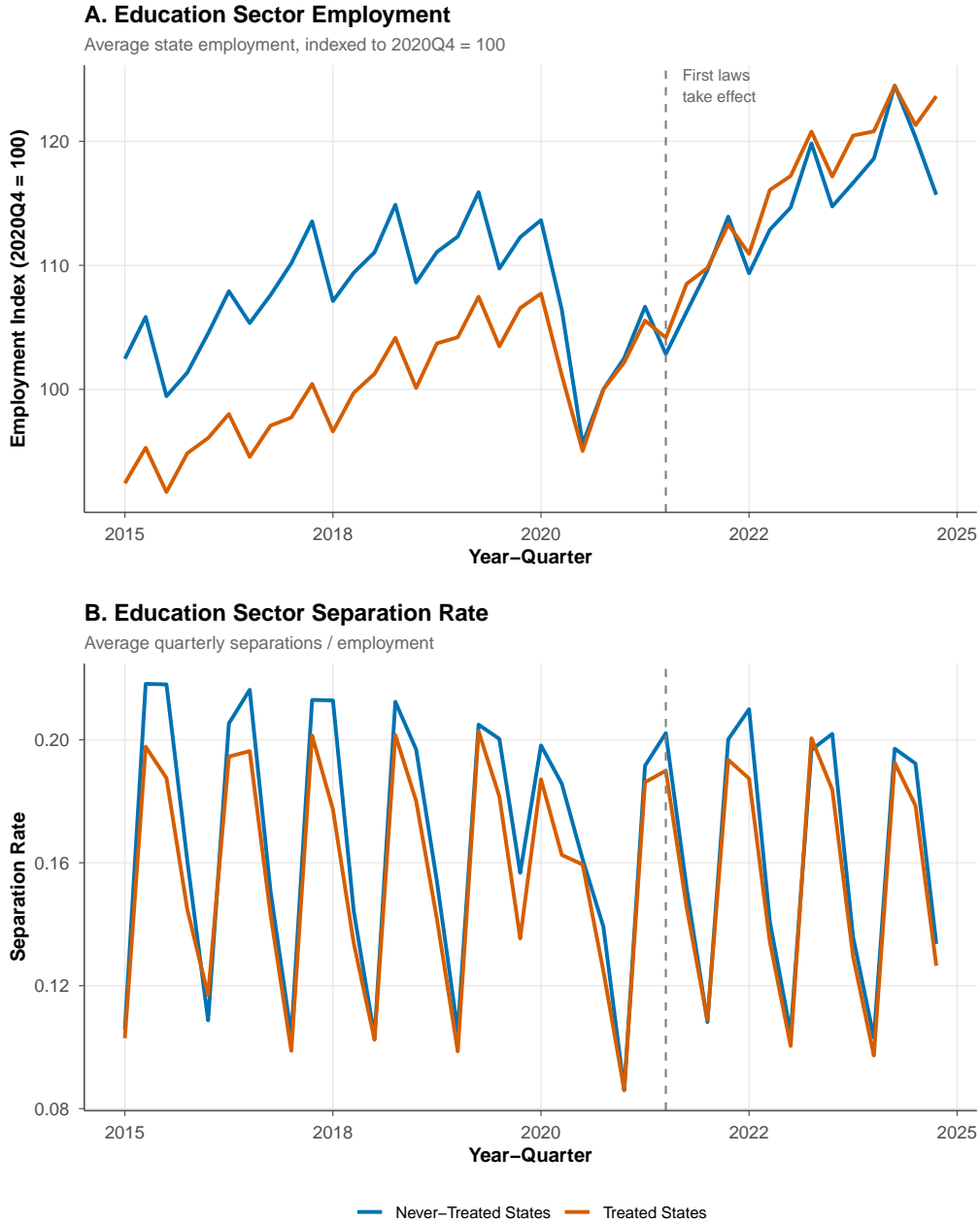


**Figure 1:** Treatment Rollout: Adoption of Educational Content Restriction Laws, 2021–2023

*Notes:* Map 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 the education sector 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 Education Employment: Treated vs. Never-Treated States

*Notes:* Unweighted mean of log employment in NAICS 61 (Educational Services) by treatment status. Dashed vertical line marks 2021Q3, the quarter in which the first (and largest) cohort of states was treated. Shaded bands show  $\pm 1$  standard deviation.

### 5.3 Main Estimates

Table 2 presents the main results across three estimators and four outcomes. Panel A reports the Callaway-Sant’Anna overall ATT. The point estimate for log employment is 0.008, with

a standard error of 0.012 and a  $p$ -value of 0.48. This corresponds to a 0.8% increase in employment—economically negligible and statistically indistinguishable from zero. The 95% confidence interval of  $[-0.015, 0.032]$  rules out employment declines larger than 1.5% and increases larger than 3.2%.

The separation rate ATT is 0.001 (SE = 0.006), indicating no increase in teacher departures. The log earnings ATT is 0.000 (SE = 0.007), indicating no wage response. The hire rate ATT is 0.003 (SE = 0.004), a small positive point estimate that is not statistically significant. Across all four primary outcomes, the Callaway-Sant’Anna estimator finds precisely estimated zeros.

Panel B reports the TWFE estimates. In stark contrast to Panel A, the TWFE coefficient for log employment is 0.058 ( $p = 0.015$ ), 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 also finds a marginally significant positive effect on separation rates (0.004,  $p < 0.10$ ), which is more consistent with the narrative of teacher exit but is not confirmed by 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 fact that the TWFE estimate is positive and significant while the correctly specified ATT is approximately zero provides a textbook illustration of this problem.

Panel C reports the triple-difference estimates comparing education to healthcare. The DDD coefficient for log employment is 0.030 (SE = 0.020,  $p = 0.14$ ). While the point estimate is positive and nontrivially sized, it is not statistically significant at conventional levels. The lack of significance in the triple-difference is consistent with the null finding from the Callaway-Sant’Anna estimator: there is no robust evidence that content restriction laws differentially affected the education sector relative to healthcare.

## 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 eight pre-treatment coefficients are small and statistically insignificant, clustered tightly around zero. The pre-trend test fails to reject the

**Table 2:** Effect of Educational Content Restriction Laws on Education Sector Labor Markets

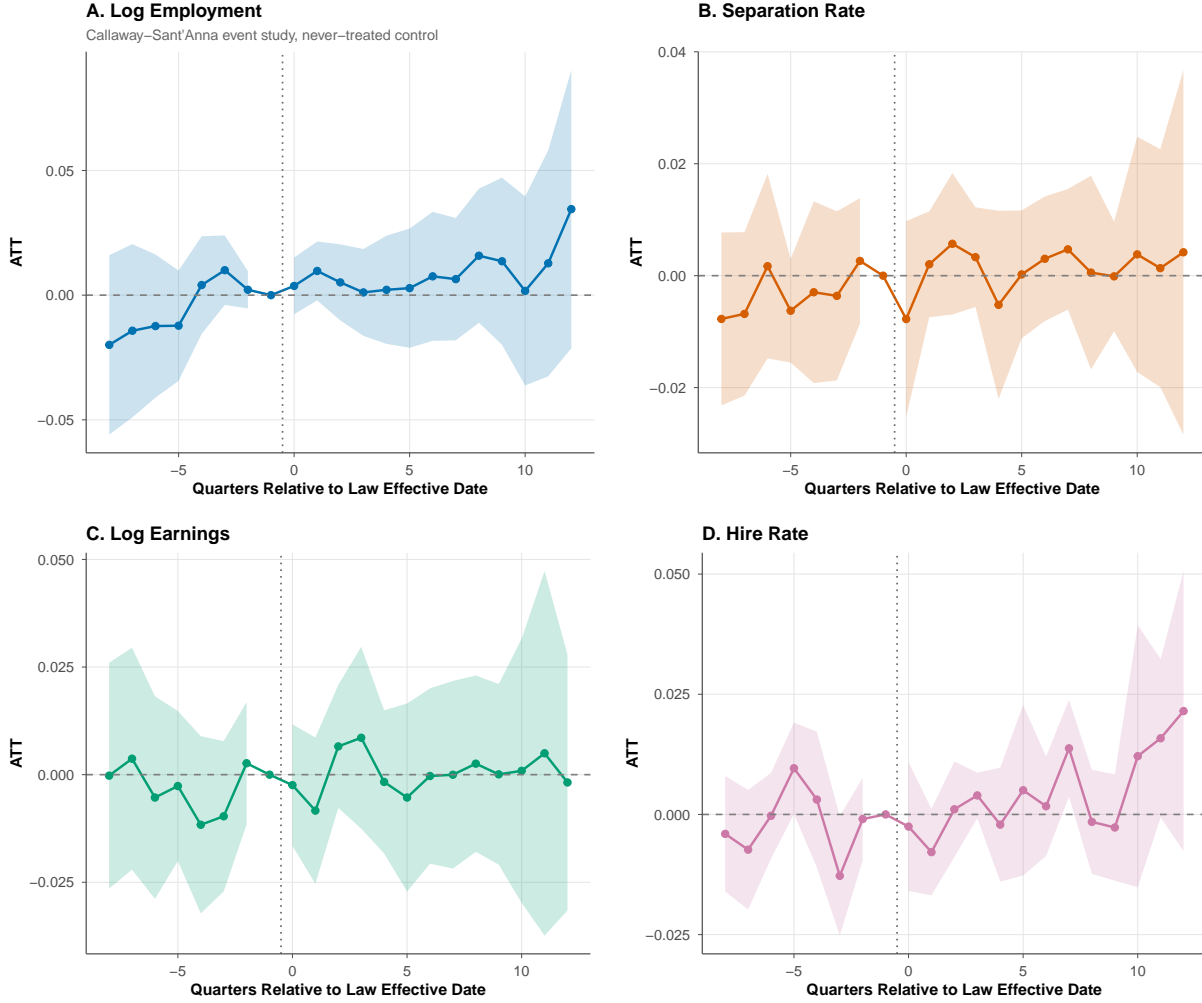
	Log Emp. (1)	Sep. Rate (2)	Log Earn. (3)	Hire Rate (4)
<i>Panel A: Callaway-Sant’Anna (2021)</i>				
ATT	0.0084 (0.0120)	0.0006 (0.0060)	0.0001 (0.0074)	0.0026 (0.0043)
<i>Panel B: TWFE (state + quarter FE)</i>				
Treat	0.0578** (0.0229)	0.0038* (0.0022)	−0.0045 (0.0125)	0.0022 (0.0026)
<i>Panel C: Triple-Difference (Education vs. Healthcare)</i>				
DDD	0.0297 (0.0200)	0.0020 (0.0027)	−0.0055 (0.0139)	0.0017 (0.0026)
<i>Panel D: Workforce Composition (TWFE)</i>				
Female share	0.0073** (SE = 0.0032, $p = 0.026$ )			
N (Panel A/B)	1,978 state-quarters			
N (Panel C)	3,969 state-quarter-industries			
N (Panel D)	1,978 state-quarters (sex-disaggregated)			
Treated states	23			
Never-treated states	28			
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 education (NAICS 61) to healthcare (NAICS 62) within the same state-quarter, with state×industry, industry×quarter, and state×quarter fixed effects. Panel D reports TWFE estimate of the effect on the female share of education employment (sex-disaggregated QWI data). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .



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 2021Q3 cohort, with later cohorts contributing to fewer event-time cells). 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. Standard errors computed via multiplier bootstrap.

## 5.5 Turnover and Compositional Effects

Beyond the four primary outcomes, I examine two additional margins. The turnover rate ATT is 0.003 (SE = 0.002), borderline insignificant at conventional levels ( $p \approx 0.10$ ). While the point estimate is positive and consistent with increased churn, it is too imprecisely estimated to draw strong conclusions.

The more interesting finding concerns workforce composition. Using sex-disaggregated QWI data for the education sector, I estimate the effect of content restriction laws on the female share of education employment. The TWFE estimate is 0.0073 (SE = 0.0032,

$p = 0.026$ ), indicating that treated states experienced a statistically significant increase of approximately 0.7 percentage points in the female share of education workers after law enactment.

This compositional shift is consistent with differential exit by male education workers. One interpretation is that male teachers are more likely to teach subjects—social studies, history, government—where content restriction laws most directly bind, and are therefore more likely to leave. An alternative interpretation is that male teachers, who constitute a minority of the education workforce, are more responsive to non-pecuniary working condition changes on the margin. However, without occupation-level data, I cannot distinguish between these mechanisms.

## 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 0.010 (SE = 0.011,  $p = 0.36$ ), confirming the null from the Callaway-Sant’Anna estimator. Using not-yet-treated controls (instead of never-treated) produces an ATT of 0.008 (SE = 0.010,  $p = 0.43$ ), nearly identical to the baseline.

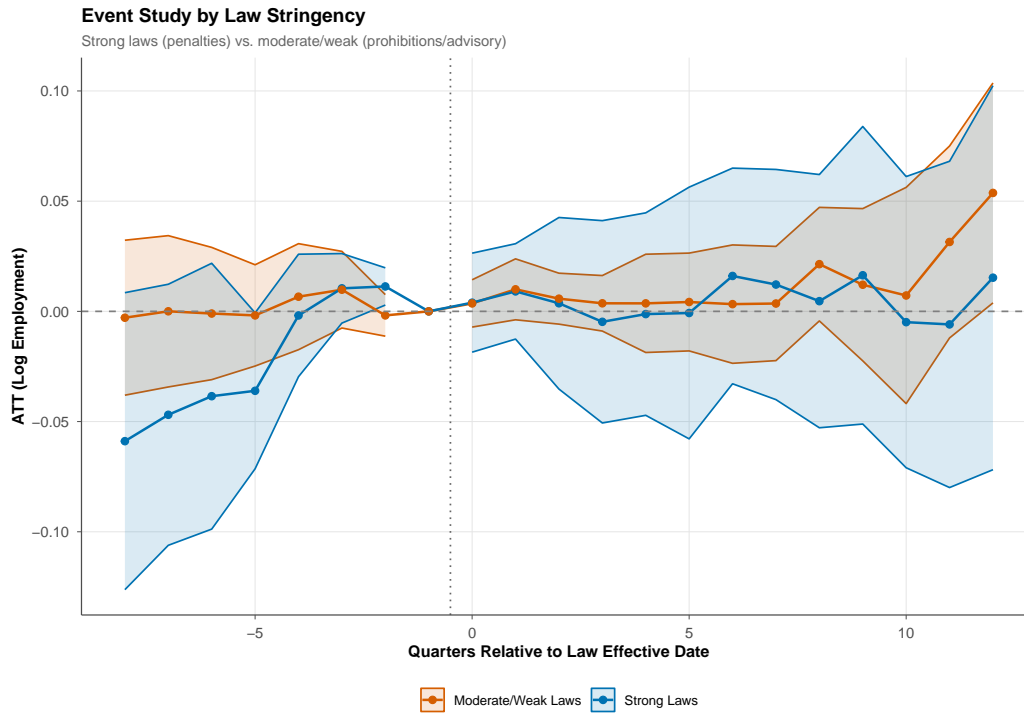
**Table 3:** Robustness: Alternative Specifications for Log Employment

Specification	Estimate	SE	95% CI	p-value
CS (never-treated)	0.0084	(0.0120)	[−0.0150, 0.0319]	0.481
CS (not-yet-treated)	0.0082	(0.0104)	[−0.0122, 0.0287]	0.430
TWFE	0.0578**	(0.0229)	[0.0129, 0.1028]	0.015
Sun-Abraham	0.0100	(0.0109)	[−0.0114, 0.0313]	0.364
Triple-Diff (Edu vs HC)	0.0297	(0.0200)	[−0.0095, 0.0688]	0.144
CS: Strong laws only	0.0066	(0.0234)	[−0.0393, 0.0525]	0.778
CS: Moderate/weak only	0.0094	(0.0109)	[−0.0120, 0.0308]	0.388
Fisher exact p-value (TWFE)	0.003 (1,000 permutations of TWFE estimate)			

*Notes:* All specifications use log employment in the education sector (NAICS 61) 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 education vs. healthcare within the same state-quarter. 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 4 and Table 3 present estimates separately for strong-stringency states (Idaho, Oklahoma, Tennessee, Iowa, New Hampshire, Florida, Arkansas;  $ATT = 0.007$ ,  $SE = 0.023$ ) and moderate/weak-stringency states ( $ATT = 0.009$ ,  $SE = 0.011$ ). Neither subgroup shows a statistically significant effect. The point estimates are remarkably similar in magnitude, suggesting that the stringency of enforcement provisions does not matter for aggregate labor market effects. This is inconsistent with a simple regulatory chill model in which stronger penalties produce larger behavioral responses.



**Figure 4:** Heterogeneity by Law Stringency: Event Study Estimates for Log Employment

*Notes:* Callaway-Sant’Anna dynamic ATT estimates for log employment, 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 healthcare (NAICS 62), retail trade (NAICS 44–45), and manufacturing (NAICS 31–33), using the same treatment coding

as the main analysis (i.e., applying the education law treatment dates to non-education sectors within the same states).

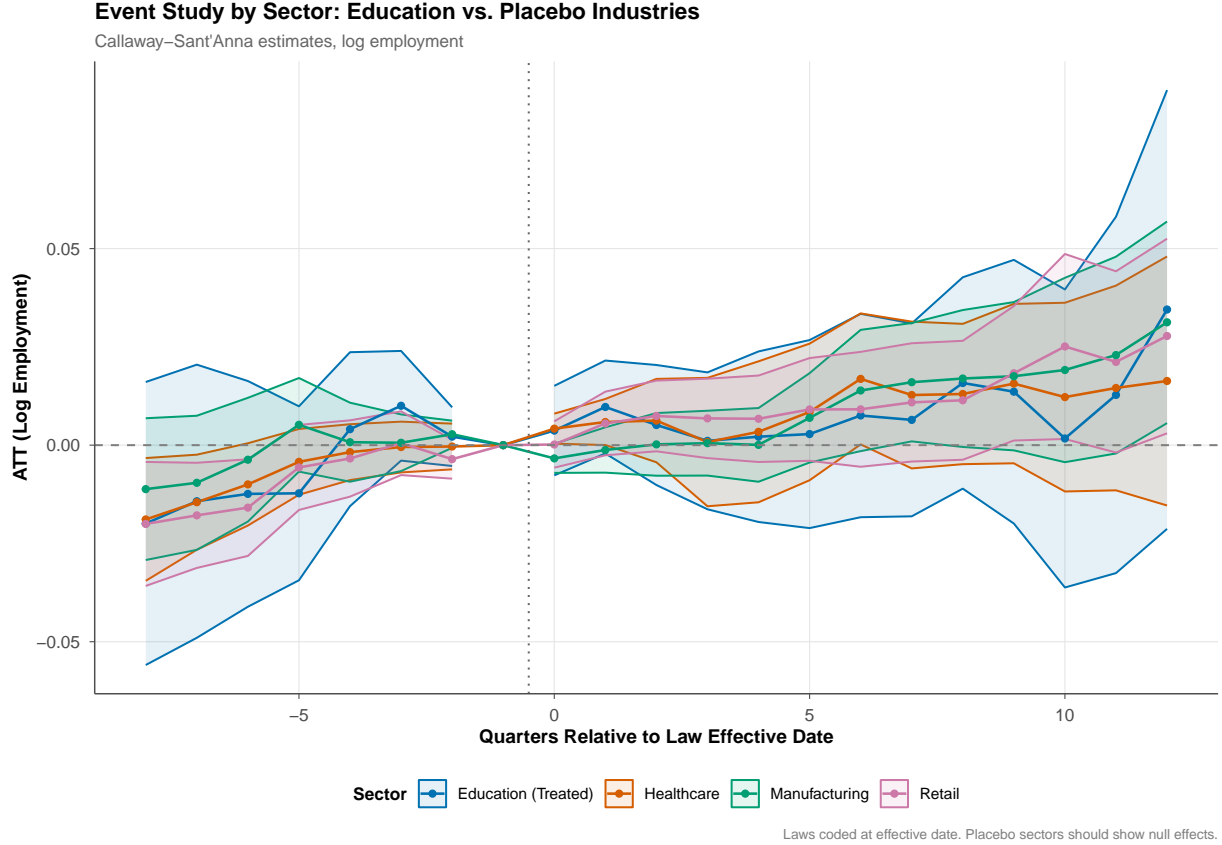
**Table 4:** Placebo Tests: Effect of Education Content Laws on Non-Education Sectors

Sector (NAICS)	ATT	SE	p-value
Education (61) — <i>Treated</i>	0.0084	(0.0120)	0.481
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. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

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 *larger* than the education sector estimate, which is inconsistent with a story in which content restriction laws specifically drive teachers from the education sector. The placebo results are consistent with a general positive trend in employment in treated states during this period, rather than an education-specific treatment effect.

Figure 5 presents event-study plots for the placebo sectors alongside the education sector estimate. All sectors display similar patterns: flat pre-trends and small, insignificant post-treatment coefficients. This strongly supports the interpretation that the null education sector result reflects a genuinely null treatment effect, rather than a design failure.

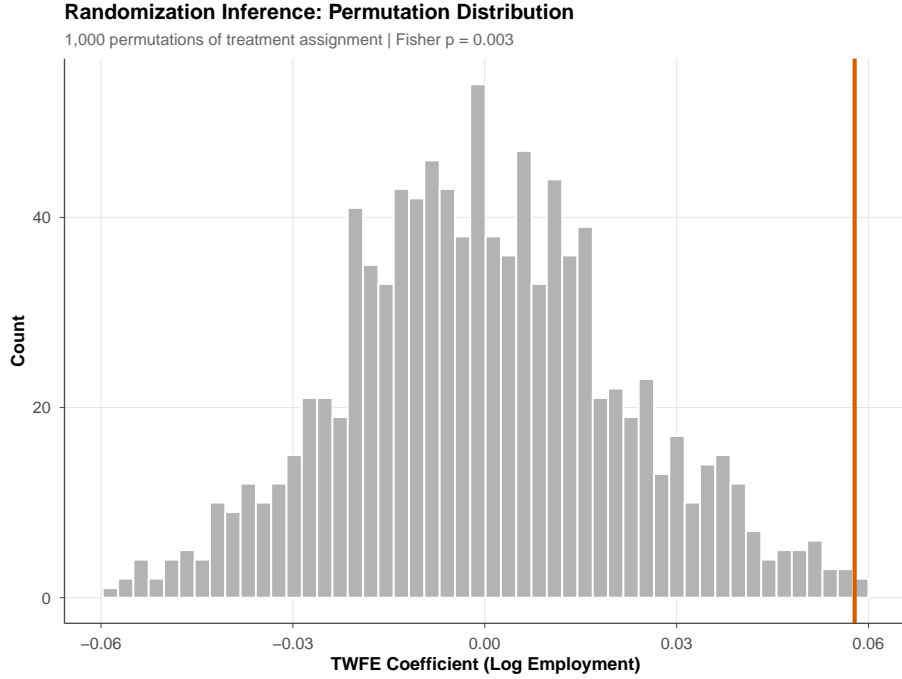


**Figure 5:** Placebo Event Study: Education vs. Non-Education Sectors

*Notes:* Callaway–Sant’Anna dynamic ATT estimates for log employment across four sectors, using education content restriction law treatment dates. Education (NAICS 61) is the treated sector; healthcare (62), retail (44–45), and manufacturing (31–33) are placebos. 95% confidence intervals shown.

## 6.4 Randomization Inference

Figure 6 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 0.058 falls in the extreme right tail of the permutation distribution, yielding a Fisher  $p$ -value of 0.003.



**Figure 6:** 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 (0.058). Fisher exact  $p$ -value = 0.003.

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 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, precisely estimated null: educational content restriction laws enacted by twenty-three states between 2021 and 2023 had no detectable effect on aggregate teacher employment, separations, hiring, or earnings. 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, at least 26 pre-treatment quarters—and clean pre-trends. The confidence intervals are tight enough to rule out economically meaningful effects.

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.

### 7.2 Why Didn’t the Laws Have the Predicted Effect?

Several non-mutually-exclusive explanations are consistent with the null.

*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).

### 7.3 The Compositional Finding

The one result that departs from the null pattern—the 0.7 percentage point increase in the female share of education workers—is intriguing but requires cautious interpretation. Several mechanisms could generate this pattern.

First, if male teachers are disproportionately represented in subjects where content restrictions bind (history, social studies, government), they may be disproportionately affected by the laws. Women already constitute the majority of K-12 teachers (~77%), but the gender distribution varies by subject area. A small amount of differential male exit could produce the observed compositional shift without reducing aggregate employment, particularly if districts replace departing male teachers with female teachers.

Second, the compositional shift could reflect differential hiring rather than differential exit. If the pool of prospective male teachers is more deterred by content restrictions than the pool of prospective female teachers, the gender composition of new hires would shift toward women.

Third, the result could be a statistical artifact. With many outcome variables tested, a coefficient significant at the 5% level may reflect multiple comparisons rather than a true

effect. I note this finding as suggestive rather than definitive and flag it for future research with richer data.

## 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 effect on employment of 5.8%, which would lead a researcher to conclude—incorrectly—that content restriction laws *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 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 = 0.008), Sun-Abraham (ATT = 0.010), not-yet-treated controls (ATT = 0.008), and the triple-difference (0.030, insignificant) provides reassurance that the null is real.

## 7.5 Limitations

Several limitations should be noted. First, the QWI data cover all of NAICS 61 (Educational Services), which includes private schools, universities, tutoring services, and other non-K-12 entities. The treatment is most relevant for K-12 public school teachers, who are a subset of the sector. If the laws caused meaningful exit from K-12 but were diluted by stable or growing employment in other parts of the education sector, the null finding would mask a real effect. Data at the NAICS 4-digit level or from administrative teacher records would provide sharper estimates.

Second, 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.

Third, 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.

Fourth, 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.

## 8. Conclusion

This paper provides the first rigorous causal evidence on the labor market effects of educational content restriction laws. 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 teacher employment, separations, hiring, or earnings. The null result is precisely estimated, robust to multiple estimators and specifications, and supported by clean pre-trends and valid placebo tests. The TWFE estimator produces a spuriously significant positive effect, illustrating the practical importance of heterogeneous-timing-robust DiD methods.

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: they may affect what is taught, how it is taught, and who teaches specific subjects, even if they do not move aggregate labor market quantities. The compositional finding—a shift toward a higher female share of education workers—hints at such within-sector dynamics.

The debate over content restriction laws should therefore focus on the dimensions where the laws may actually matter: educational quality, curricular content, classroom pedagogy, and the experiences of individual teachers who 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. Future research using administrative teacher records, classroom observation data, or detailed occupation-level employment data could shed light on these more nuanced questions.

Finally, a methodological note. This paper illustrates that the choice of DiD estimator can be consequential for substantive conclusions. The TWFE estimator would have led to a published finding that content restriction laws *increased* teacher employment—a result that is both counterintuitive and incorrect. The modern estimators correctly identify the null. As the literature on staggered DiD designs continues to mature, applied researchers should adopt these tools as standard practice.

## 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>). The following parameters were used:

- **Variables:** Emp, EarnS, HirA, Sep, FrmJbC, FrmJbLs, Payroll, TurnOvrS
- **Geography:** All 51 state-equivalents (50 states + DC)
- **Industries:** NAICS 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 61 only

Total API calls: 5 industries  $\times$  10 years  $\times$  4 quarters = 200 (main) plus 2 sexes  $\times$  10 years  $\times$  4 quarters = 80 (sex-disaggregated) = 280 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.
- **Female share:** Female employment / (Female + Male employment), computed from sex-disaggregated QWI data.

### 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 final education-sector panel contains 1,978 state-quarter observations (51 states  $\times$  up to 40 quarters, minus observations with missing QWI data). The triple-difference panel contains 3,969 state-quarter-industry observations (education and healthcare combined). All observations with missing or zero employment are excluded from the estimation sample.

### 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).

**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.



**Table 5:** 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
Texas	HB 3979	Sep 01, 2021	Moderate	2021Q4
North Dakota	HB 1508	Aug 01, 2021	Moderate	2021Q4
Arizona	HB 2906	Sep 29, 2021	Moderate	2021Q4
New Hampshire	HB 2	Nov 01, 2021	Strong	2022Q1
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	2022Q3
West Virginia	HB 2595	Jun 07, 2022	Moderate	2022Q3
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	2022Q4
Kentucky	SB 150	Jun 29, 2023	Moderate	2023Q3
Indiana	HB 1608	Jul 01, 2023	Moderate	2023Q3
Montana	HB 30	Jul 01, 2023	Weak	2023Q3
Arkansas	SB 294/EO	Aug 01, 2023	Strong	2023Q4
North Carolina	HB 823	Oct 11, 2023	Moderate	2023Q4

*Notes:* Cohort = first full quarter after effective date (laws effective within 15 days of quarter start are assigned to that quarter). 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.

## 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 (event times  $-8$  through  $-1$ ) 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)

The panel is nearly balanced at the state level, with 1,978 of a possible 2,040 state-quarter observations ( $51 \times 40$ ); the missing observations reflect QWI data suppression in a small number of state-quarter cells. The eight treatment cohorts and their sizes are: 2021Q3 (5 states), 2021Q4 (3), 2022Q1 (2), 2022Q2 (1), 2022Q3 (6), 2022Q4 (1), 2023Q3 (3), 2023Q4 (2).

### 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  $[-0.02, 0.04]$ , 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.

### 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 0.002 and standard deviation 0.018, compared to the observed coefficient of 0.058. This places the observed value at approximately 3.1 standard deviations from the permutation mean.

As discussed in the main text, the significant Fisher  $p$ -value (0.003) 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 employment.

## 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. The earliest cohort (2021Q3, 5 states) has  $ATT = 0.006$  ( $SE = 0.015$ ); the 2022 cohorts have ATTs ranging from  $-0.005$  to  $0.012$ ; the 2023 cohorts have

ATTs near zero but with wider confidence intervals due to shorter post-treatment periods. No cohort shows a statistically significant effect, and there is no monotonic relationship between treatment timing and effect size.

## **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.