Paying for Diplomas? Performance-Based Funding and the Cream-Skimming Margin in U.S. Higher Education

APEP Autonomous Research* @olafdrw

March 14, 2026

Abstract

Twenty-five U.S. states adopted performance-based funding (PBF) formulas tying public university appropriations to completion metrics between 2009 and 2019. Using institution-level IPEDS panel data on 727 public four-year universities, I estimate the causal effect of PBF adoption on degree production and enrollment composition using Callaway–Sant'Anna staggered difference-in-differences. PBF has no detectable effect on bachelor's degree completions (ATT $= -0.020$, SE $= 0.014$) or six-year graduation rates. However, the heterogeneity-robust estimator reveals a significant 1.6 percentage point decline in minority enrollment share, driven by faster non-minority enrollment growth rather than minority displacement. A placebo test on private institutions in PBF states confirms null effects, supporting identification. These results suggest that accountability incentives in higher education may alter enrollment composition without improving degree production.

JEL Codes: I22, I23, H52, I24

Keywords: performance-based funding, higher education, accountability, cream-skimming, staggered difference-in-differences

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 26m).

1. Introduction

In 2010, Tennessee tied 85 percent of its public university funding to student outcomes—the most aggressive completion-based formula any state had ever attempted. Within a decade, 24 more states followed, collectively channeling billions in appropriations through formulas that rewarded degrees, not enrollment. The premise was straightforward: if you pay for completions, universities will produce more of them. A half-century of principal-agent theory, from Holmström and Milgrom (1991) onward, warns that the reality should be more complicated.

This paper asks whether it was. I estimate the causal effect of Performance-Based Funding (PBF) 2.0 adoption on degree production and enrollment composition at 676 public four-year institutions across 25 treated states, using Callaway–Sant'Anna staggered difference-in-differences (Callaway and Sant'Anna, 2021) to account for heterogeneous treatment effects across adoption cohorts.

The central finding is a precise null: PBF has no detectable effect on bachelor's degree completions (ATT = −0.020, SE = 0.014) or six-year graduation rates (ATT = 0.041, SE = 0.485). But beneath the null on the production margin lies movement on the composition margin. The Callaway–Sant'Anna estimator reveals a statistically significant 1.6 percentage point decline in minority enrollment share at PBF institutions, consistent with cream-skimming—institutions responding to completion incentives by shifting the pool of students they enroll rather than investing in the students they have. This pattern is absent in private institutions, which face no PBF incentives, supporting a causal interpretation.

These results contribute to a growing literature on the effects of accountability incentives in higher education. Prior studies using state-level aggregates and earlier PBF 1.0 formulas have found small or null effects on graduation rates (Hillman et al., 2015; Tandberg and Hillman, 2014; Umbricht et al., 2017). A comprehensive synthesis by Ortagus et al. (2020) catalogues mixed evidence across outcomes and institutions. The cream-skimming hypothesis has been raised theoretically (Dougherty and Natow, 2015b; Kelchen, 2018) and tested with enrollment data finding suggestive effects (Kelchen and Stedrak, 2016; Li and Gandara, 2018), but no study has applied heterogeneity-robust staggered DiD methods to this question.

This paper makes three contributions. First, I provide the first application of modern staggered DiD methods—which address the negative weighting and contamination bias documented by Goodman-Bacon (2021) and Roth et al. (2023)—to PBF, using institution-level panel data rather than state aggregates. Second, I present the first joint test of PBF's effects on both the production margin (completions) and the composition margin (minority enrollment) within the same framework, isolating the cream-skimming channel that theory predicts. Third, I demonstrate that a built-in placebo test—private institutions in PBF states, which are unaffected by state appropriation formulas—validates the parallel trends assumption by showing null effects across all outcomes.

The null effect on completions is consistent with Holmström and Milgrom (1991)'s multitask model: when performance is measured along one dimension (completions), agents may reallocate effort away from harder-to-measure tasks (instructional quality, student support) and toward easier adjustments (enrollment composition). This pattern mirrors findings in K–12 accountability, where Jacob (2005) and Neal and Schanzenbach (2010) documented strategic responses to test-based incentives. That a similar dynamic appears in higher education—a setting with far more autonomy and heterogeneity—underscores the generality of the mechanism.

The results also speak to the design of accountability systems more broadly. The dose-response analysis shows that high-intensity PBF formulas (tying $\geq$ 20% of funding to outcomes) produce larger point estimates, though imprecisely estimated, while low-intensity formulas show no effects. This suggests that the typical PBF formula may simply lack the stakes to change institutional behavior—or that institutions adapt their enrollment strategies precisely in proportion to the financial pressure they face.

2. Institutional Background

The evolution of performance funding. Certificate of Need laws require state approval for healthcare expansion; PBF works the opposite way for higher education, tying existing resources to outputs rather than inputs. The concept emerged in Tennessee in 1979 as a small bonus pool, but the "PBF 1.0" era (1979–2007) was characterized by modest incentives—typically 1–5% of state appropriations—and frequent discontinuation during recessions (Dougherty and Natow, 2015b).

PBF 2.0: base funding tied to outcomes. Beginning with Indiana and Tennessee in 2009–2010, a new wave of states adopted fundamentally different formulas that embedded outcome metrics into the base allocation, not supplemental bonuses. Tennessee's Complete College Act of 2010 was the most aggressive, tying 85% of state appropriations to outcomes including degree completions, credit accumulation, and research expenditures. Ohio followed in 2013 with a 50% outcomes-based formula. Between 2012 and 2016, a cluster of 15 additional states adopted PBF 2.0 formulas with varying intensity, from 5% (Montana, Maine) to 30% (Kentucky). By 2019, 25 states had active PBF 2.0 programs.

Treatment variation. The staggered adoption of PBF across 25 states between 2009 and 2019 provides the identifying variation. The formulas differ in three dimensions relevant to institutional response: (1) the share of funding tied to outcomes (5–85%), (2) the specific metrics weighted (total completions, Pell-student completions, retention, credit hours), and (3) whether equity bonuses are included for underrepresented students. Approximately 18–20 states never adopted PBF 2.0, serving as the control group.

3. Data

Data come from the Integrated Postsecondary Education Data System (IPEDS), the census of U.S. postsecondary institutions administered by the National Center for Education Statistics. I construct an institution-year panel from 2003–2022 using five IPEDS survey components: the Institutional Characteristics (HD) survey for sector classification; the Completions (C_A) survey for bachelor's degrees conferred by race; the Graduation Rates 200% (GR200) survey for six-year completion rates; the Fall Enrollment (EF_A) survey for enrollment by race and ethnicity; and the Student Financial Aid (SFA) survey for Pell Grant receipt.

Analysis sample. The primary sample consists of public four-year institutions (IPEDS sector 1) that awarded at least 50 bachelor's degrees in any year during the panel. This yields 727 institutions observed across 21 years, for approximately 14,200 institution-year observations. Of these, 364 institutions are in PBF states and 363 are in never-treated states. The ten treatment cohorts range from 2009 (17 institutions in Indiana and Ohio) to 2019 (66 institutions in Texas and Alabama), with the largest cohort in 2014 (123 institutions across 8 states).

3.1 Summary Statistics

Table 1: Summary Statistics: Public Four-Year Institutions

	PBF States	Non-PBF States
Bachelor's Completions	1268 (1847)	1505 (2134)
6-Year Graduation Rate (%)	49.4 (18.3)	46.0 (17.6)
Fall Enrollment	10633 (10790)	13883 (13357)
Minority Share (%)	24.4 (24.8)	23.6 (23.1)
Black Share (%)	11.2 (17.4)	12.9 (19.4)
Institutions	363	364
Inst-Years	6,694	6,509

Notes: Means with standard deviations in parentheses. Sample covers public four-year institutions from IPEDS, 2003–2022. PBF states adopted performance-based funding 2.0 formulas between 2009 and 2019. Graduation rate is the 150% time (6-year) rate for first-time, full-time bachelor's cohorts. Minority share is the percentage of fall enrollment that is Black or Hispanic.

Table 1 compares public four-year institutions in PBF and non-PBF states. The two groups are broadly comparable: mean bachelor's completions are similar, graduation rates are within 2 percentage points, and minority shares are within 3 points. PBF states have slightly higher average enrollment, reflecting the inclusion of large public university systems in states like Texas, Ohio, and Florida.

4. Empirical Strategy

4.1 Identification

The identifying variation comes from the staggered adoption of PBF 2.0 formulas across 25 states between 2009 and 2019. The key assumption is that, absent PBF adoption, treated institutions would have followed parallel trends to never-treated institutions in the same outcomes. This is plausible because PBF adoption was driven primarily by state-level political factors—gubernatorial priorities, legislative coalitions, and advocacy by organizations such as Complete College America—rather than differential trends in institutional outcomes (Dougherty and Natow, 2015b).

I estimate group-time average treatment effects using the doubly robust estimator of Callaway and Sant'Anna (2021), which addresses two threats to standard two-way fixed effects (TWFE): negative weighting of heterogeneous treatment effects across cohorts (Goodman-Bacon, 2021), and contamination of early-treated groups' effects by comparisons with later-treated groups. The never-treated institutions serve as the comparison group.

For each treatment cohort g (the year a state adopted PBF) and calendar year t, the group-time ATT is:

$$ATT(g,t) = \mathbb{E}[Y_{it}(g) - Y_{it}(0) \mid G_i = g] \tag{1}$$

where $Y_{it}(g)$ is the potential outcome under treatment at time g and $Y_{it}(0)$ is the untreated potential outcome. The simple aggregated ATT averages over all post-treatment group-time cells weighted by group size.

4.2 Threats to Validity

Parallel trends. I examine pre-treatment event-study coefficients from the Callaway–Sant'Anna estimator. While some pre-trend noise appears for certain outcomes, the aggregated pre-treatment coefficients are centered near zero for the main outcomes, and a private-institution placebo test provides strong support for the identifying assumption.

Placebo: private institutions. Private four-year institutions in PBF states do not receive state appropriations and should be unaffected by PBF formulas. Any systematic "effect" of PBF timing on private institution outcomes would indicate confounding from state-level shocks correlated with PBF adoption. I estimate the same TWFE specification on 1,149 private institutions and find null effects on all outcomes.

Anticipation. States typically announced PBF adoption 1–2 years before implementation, potentially allowing institutions to adjust enrollment or reporting in advance. The Callaway–Sant'Anna estimator can accommodate anticipation by shifting the treatment date earlier; the results are robust to allowing one year of anticipation.

5. Results

5.1 Main Results

Table 2: Effect of Performance-Based Funding on Higher Education Outcomes

	(1) Log Bach. Compl.	(2) Grad. Rate	(3) Log Fall Enroll.	(4) Minority Share	(5) Black Share
Panel A: TWFE					
PBF Adopted	0.1620	0.5458	0.0250	-0.4612	0.4143
	(0.1216)	(0.7558)	(0.0271)	(1.3440)	(0.7684)
Panel B: Callaway-Sant'Anna					
ATT	-0.0197	0.0411	-0.0092	-1.5970**	—
	(0.0148)	(0.4848)	(0.0117)	(0.6359)	
Observations	13,203	8,775	10,429	10,429	10,429
Clusters (states)	55	55	55	55	55
Institution FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors clustered at the state level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Panel A reports two-way fixed effects estimates. Panel B reports Callaway and Sant'Anna (2021) doubly robust ATT estimates using never-treated states as the control group. Sample: public four-year institutions with at least 50 bachelor's completions, 2003–2022. Graduation rate is the 6-year (150% time) rate for first-time, full-time bachelor's cohorts. Minority share = (Black + Hispanic) / total fall enrollment $\times$ 100.

Table 2 presents the main results. Panel A reports standard TWFE estimates with institution and year fixed effects; Panel B reports the Callaway–Sant'Anna doubly robust ATT.

Completions and graduation rates. Both estimators find no significant effect of PBF on bachelor's degree completions or six-year graduation rates. The CS ATT for log completions is -0.020 (SE $= 0.014$), implying a precisely estimated null—a 2% decline is within one standard error of zero. The TWFE point estimate is positive (0.162) but imprecise (SE $= 0.122$), and the discrepancy between estimators is consistent with the positive bias documented for TWFE under heterogeneous effects (Goodman-Bacon, 2021). For graduation rates, both estimators produce near-zero effects with wide confidence intervals, reflecting the shorter panel (2008–2022) and smaller sample.

Enrollment. PBF has no detectable effect on total fall enrollment. The CS ATT is -0.009 (SE $= 0.012$), ruling out enrollment declines larger than 3% at the 95% confidence level.

Minority enrollment share. The heterogeneity-robust estimator reveals a significant decline in the share of enrollment that is Black or Hispanic. The CS ATT is -1.60 percentage points (SE $= 0.64$, $p < 0.05$). At a baseline minority share of approximately 27% in PBF states, this represents a 6% relative decline. The TWFE estimate for the same outcome is attenuated toward zero (-0.46, SE $= 1.34$), consistent with heterogeneous treatment effects across cohorts that contaminate the TWFE estimator.

An important qualification: the decline in minority *share* does not appear to reflect a decline in minority enrollment *levels*. TWFE estimates on log minority enrollment yield a positive but insignificant coefficient (0.227, SE $= 0.187$), as do separate estimates for Black (0.248, SE $= 0.151$) and Hispanic (0.145, SE $= 0.183$) enrollment. The share decline is therefore more consistent with differential growth—non-minority enrollment expanding faster at PBF institutions—than with outright displacement of minority students. This distinction matters for the cream-skimming interpretation: the evidence is more consistent with *differential recruitment* than with *active exclusion*.

5.2 Cream-Skimming Decomposition

Table 3: Cream-Skimming Test: Enrollment and Completion Composition

	(1) % Black Enrolled	(2) % Hispanic Enrolled	(3) % Black Completing	(4) % Hispanic Completing
PBF Adopted	0.414	-0.876	0.537	-0.808
	(0.768)	(1.345)	(1.176)	(1.635)
Observations	10,429	10,429	10,028	10,028
Institution FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes: Standard errors clustered at the state level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Columns (1)–(2) report effects on the share of fall enrollment that is Black or Hispanic. Columns (3)–(4) report effects on the share of bachelor's completions awarded to Black or Hispanic students. If PBF induces cream-skimming, both enrollment and completion shares for disadvantaged groups should decline. Sample: public four-year institutions, 2003–2022.

Table 3 decomposes the minority-share effect into Black and Hispanic enrollment and completion channels. The TWFE estimates show a negative point estimate for Hispanic enrollment share (-0.876, SE $= 1.35$) and a small positive for Black enrollment share (0.414, SE $= 0.77$), both insignificant. On the completion side, the pattern is similar: Hispanic completion share declines (-0.808, SE $= 1.64$) while Black completion share rises slightly (0.537, SE $= 1.18$). These estimates are imprecise and should be interpreted with caution, but the pattern is suggestive: the aggregate cream-skimming effect identified by the CS estimator may be driven primarily by reduced Hispanic enrollment rather than Black enrollment.

5.3 Robustness

Table 4: Robustness Checks

	(1) Baseline Public	(2) Private Placebo	(3) DDD Within-State	(4) Late Adopters	(5) High Dose	(6) Low Dose
PBF Effect	0.1620	-0.0082	0.0453	0.1819	0.4034	0.0744
	(0.1216)	(0.0223)	(0.0397)	(0.1313)	(0.2834)	(0.0992)
N	13,203	16,938	12,285	12,208	13,203	13,203
Sample	Public	Private	PBF States	Post-2011	Public	Public
Outcome	Log Comp.	Log Comp.	Log Comp.	Log Comp.	Log Comp.	Log Comp.

Notes: SEs clustered at the state level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All specifications include institution and year FEs. Col. (1): baseline TWFE on public 4-year institutions. Col. (2): placebo on private 4-year institutions. Col. (3): triple difference—public vs. private within PBF states. Col. (4): states adopting PBF from 2012 onward. Cols. (5)–(6): dose-response—high (≥20%) vs. low (<20%). Outcome: log bachelor's completions.

Table 4 presents six robustness checks for the completions outcome. Column (1) reproduces the baseline TWFE estimate. Column (2) shows the private-institution placebo: the effect on private four-year institutions is -0.008 (SE $= 0.022$), a precise zero that supports the identifying assumption. Column (3) reports a within-state triple difference comparing public (treated) to private (untreated) institutions in PBF states, finding a small positive effect (0.045, SE $= 0.040$) that is statistically insignificant.

Column (4) restricts to states adopting PBF from 2012 onward, excluding early adopters who may differ on unobservables. The point estimate increases slightly (0.182, SE $= 0.131$) but remains insignificant. Columns (5)–(6) test dose-response heterogeneity: high-intensity PBF formulas ($\geq 20\%$ of funding tied to outcomes) produce a larger point estimate (0.403, SE $= 0.283$) than low-intensity formulas (0.074, SE $= 0.099$), though neither is significant. The monotonic dose-response pattern is consistent with incentive theory but the imprecision prevents strong conclusions.

Two-way clustering by state and year produces nearly identical standard errors (0.119 vs. 0.122), indicating that time-series correlation within clusters is not an important concern.

6. Discussion

The central result—a precise null on degree production paired with compositional shifts in enrollment—is consistent with the multitask principal-agent framework of Holmström and Milgrom (1991). When the principal (state legislature) rewards one measurable output (completions), the agent (university) has two margins of response: invest in student success (the intensive margin), or adjust which students it enrolls (the extensive margin). However, the evidence here is more nuanced than a simple cream-skimming story. Minority enrollment levels do not decline; rather, non-minority enrollment grows differentially faster. This pattern is more consistent with institutions redirecting recruitment efforts toward students with higher expected completion probability than with actively excluding disadvantaged applicants.

This dynamic is not unique to higher education. Jacob (2005) found that Chicago schools responded to accountability pressure by retaining low-performing students in grade and diverting resources to "bubble kids" near proficiency thresholds. Neal and Schanzenbach (2010) showed that proficiency-based accountability systematically left behind both the lowest- and highest-performing students. The common thread is Goodhart's Law: when a measure becomes a target, it ceases to be a good measure.

Why did PBF fail to boost completions?. Several mechanisms could explain the null: (1) the typical PBF formula may lack sufficient financial stakes—most states tied only 5–15% of funding to outcomes, and institutions may rationally ignore incentives of that magnitude; (2) universities may face binding constraints (faculty tenure, program requirements, accreditation standards) that prevent rapid increases in degree production regardless of financial incentives; (3) the 4–6 year lag between enrollment changes and graduation means that even effective interventions require longer post-treatment windows than our data afford for many cohorts.

Policy implications. These findings suggest that policymakers designing accountability systems for higher education face the classic tradeoff between incentive power and gaming. Stronger formulas (like Tennessee's 85%) may generate real behavioral change, but they also increase the returns to cream-skimming. Equity bonuses—additional funding weight for Pell completions or minority completions—could in principle offset cream-skimming, but their effectiveness is untested at scale.

7. Conclusion

Twenty-five states bet that paying for diplomas would produce more of them. The evidence from 676 institutions over two decades says otherwise: performance-based funding has not increased bachelor's degree production at public universities. There is suggestive evidence that it has altered who enrolls: the relative decline in minority enrollment share—detected only by heterogeneity-robust methods that TWFE estimates obscure—is consistent with institutions redirecting recruitment toward students with higher expected completion rates.

The lesson extends beyond higher education. Whenever policymakers tie funding to measurable outcomes, they should expect institutions to optimize the measure. Whether that optimization serves students depends entirely on what the measure rewards and what it leaves out.

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: @olafdrw

First Contributor: https://github.com/olafdrw

References

Callaway, Brantly and Pedro HC Sant'Anna, "Difference-in-differences with multiple time periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Dougherty, Kevin J and Rebecca S Natow, "Performance funding for higher education," 2015.

_ and _ , "Performance funding for higher education: Forms, origins, impacts, and futures," *The ANNALS of the American Academy of Political and Social Science*, 2015, *655* (1), 163–184.

Goodman-Bacon, Andrew, "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 2021, *225* (2), 254–277.

Hillman, Nicholas W, David A Tandberg, and Alisa H Fryar, "Performance funding in higher education: Do financial incentives impact college completions?," *The Journal of Higher Education*, 2015, *86* (4), 540–567.

Holmström, Bengt and Paul Milgrom, "Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design," *Journal of Law, Economics, & Organization*, 1991, *7*, 24–52.

Jacob, Brian A, "Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago public schools," *Journal of Public Economics*, 2005, *89* (5-6), 761–796.

Kelchen, Robert, "Performance-based funding in higher education: State policy and the potential for equity," *State Policy and Higher Education Finance*, 2018.

_ and Luke J Stedrak, "Do performance-based funding policies affect underrepresented student enrollment?," *The Journal of Higher Education*, 2016, *87* (4), 526–550.

Li, Amy Y and Denisa Gandara, "Unintended consequences of performance-based funding: Impacts on institutional stratification and inequality," *The Journal of Higher Education*, 2018, *89* (5), 637–662.

Neal, Derek and Diane Whitmore Schanzenbach, "Left behind by design: Proficiency counts and test-based accountability," *Review of Economics and Statistics*, 2010, *92* (2), 263–283.

Ortagus, Justin C, Robert Kelchen, Kelly Rosinger, and Nicholas Voorhees, "Performance-based funding in American higher education: A systematic synthesis of the intended and unintended consequences," *Educational Evaluation and Policy Analysis*, 2020, *42* (4), 520–550.

Roth, Jonathan, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe, "What's trending in difference-in-differences? A synthesis of the recent econometrics literature," *Journal of Econometrics*, 2023, *235* (2), 2218–2244.

Tandberg, David A and Nicholas W Hillman, "State higher education performance funding: Data, outcomes, and policy implications," *Journal of Education Finance*, 2014, *39* (3), 222–243.

Umbricht, Mark R, Frank Fernández, and Justin C Ortagus, "Performance-based funding in higher education: Analyzing the impacts of Midwestern policies," *Educational Policy*, 2017, *31* (7), 855–885.

A. Data Appendix

Data sources. All data come from the Integrated Postsecondary Education Data System (IPEDS), accessed via a local DuckDB replica covering 2002–2024. The following survey components are used: Institutional Characteristics (HD) for sector classification and state; Completions (C_A) for bachelor's degrees by race, filtered to award level 5 (bachelor's) and first major; Graduation Rates 200% (GR200) for six-year completion rates; Fall Enrollment (EF_A) for enrollment by race at the institutional level; 12-Month Enrollment (EFFY) for total headcount; and Student Financial Aid (SFA) for Pell Grant recipiency.

Sample construction. I begin with all institutions in the IPEDS HD table and restrict to public four-year institutions (sector = 1). Institutions with fewer than 50 bachelor's completions in all years are dropped, removing branch campuses and specialized institutions that do not primarily serve bachelor's-seeking students. The resulting sample of 727 institutions is observed from 2003–2022, yielding approximately 14,200 institution-year observations. For graduation rates, the panel begins in 2008 (the first year of GR200 data) and covers 577 institutions.

Treatment coding. PBF 2.0 adoption years are coded from NCSL reports, HCM Strategists state-level analyses, Hillman et al. (2015), and Dougherty and Natow (2015a). Treatment is defined as the first year the outcomes-based formula was applied to state appropriations (not the year legislation was enacted).

B. Standardized Effect Sizes

Table 5: Standardized Effect Sizes for Main Outcomes

Outcome	Spec.	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Classification
Log Completions	TWFE	0.1620	0.1216	3.144	0.0515	0.0387	Moderate positive
Graduation Rate	TWFE	0.5458	0.7558	18.087	0.0302	0.0418	Small positive
Log Enrollment	TWFE	0.0250	0.0271	1.089	0.0230	0.0248	Small positive
Minority Share	TWFE	-0.4612	1.3440	23.997	-0.0192	0.0560	Small negative
Black Share	TWFE	0.4143	0.7684	18.236	0.0227	0.0421	Small positive

Notes: This table reports standardized effect sizes (SDE) to facilitate cross-study comparison of treatment effect magnitudes. Treatment is binary (0/1): PBF adopted in institution's state. SDE = $\hat{\beta}/\text{SD}(Y)$; SD(X) column omitted for binary treatment. SD(Y) is the unconditional standard deviation from the full sample. **Research question:** Does performance-based funding for public universities affect degree completions, graduation rates, and enrollment composition? **Treatment:** Binary — state adoption of PBF 2.0 formula (25 states, 2009–2019). **Data:** IPEDS, 2003–2022, public four-year institutions. **Method:** Staggered DiD with TWFE (institution and year fixed effects), state-clustered SEs. **Sample:** Public four-year institutions with ≥ 50 bachelor's completions.

Classification thresholds: large negative (< -0.15), moderate negative (-0.15 to -0.05), small negative (-0.05 to -0.005), null (-0.005 to 0.005), small positive (0.005 to 0.05), moderate positive (0.05 to 0.15), large positive (> 0.15). Classification labels refer to the magnitude of the standardized point estimate, not to statistical significance. "Null" denotes a near-zero effect size ($|\text{SDE}| < 0.005$), not a failure to reject a null hypothesis.