DISSERTATION

ESSAYS ON THE ECONOMICS OF EDUCATION

Submitted by

Prasiddha Shakya

Department of Economics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2025

Doctoral Committee:

Advisor: Ray Miller
Co-Advisor: Sammy Zahran

Anita Alves Pena
Nicholas Magnan

ABSTRACT

ESSAYS ON THE ECONOMICS OF EDUCATION

This dissertation comprises of three empirical studies that investigate how educational policies and disruptions influence student outcomes across K–12 and higher education contexts in the United States. The first chapter examines the impact of New Jersey's "$50K The First Day" policy, which established a $50,000 minimum salary for new teachers. Using school-level panel data and a staggered difference-in-differences design under a continuous treatment framework, I show that this salary floor increased teacher pay across the board and led to measurable improvements in student academic performance, including proficiency gains in mathematics and English Language (ELA), as well as modest increases in graduation and college enrollment rates.

The second chapter turns to postsecondary education and evaluates the Bridge Scholars Program (BSP), a pre-college summer initiative aimed at supporting first-generation and low-income students at a large public land-grant university in the United States. Employing a fuzzy regression discontinuity design within a local randomization framework, I find that BSP increases first-year and second-year persistence, with suggestive evidence that non-financial program components—such as mentoring and peer support—are central to its effectiveness.

The final chapter explores how the COVID-19 pandemic, and the associated transition to remote instruction, affected undergraduates' retention in and transition into STEM fields. Employing a staggered difference-in-differences approach with student-level longitudinal data, I explore both persistence within STEM majors and switching behavior into and out of STEM and Biology/Health related majors. Contrary to early concerns, I find little evidence of significant or lasting disruption to students' field-of-study decisions.

Taken together, these studies contribute to ongoing policy conversations around teacher compensation, college access and retention, and the postsecondary consequences of the pandemic– a major educational disruption. Each chapter offers evidence-based insights that I hope can be used for designing interventions that promote student success and equity.

ACKNOWLEDGEMENTS

Out of all the pages in this dissertation, this has been the hardest for me to write. I was very hesitant to write this page, but I decided to include it–more for future me to read and reflect back on than for anybody else.

Rojina, thanks for making the first two years of graduate school a walk in the park. I will forever be grateful to you for deciding to transfer and start your PhD in Colorado. Every day I spent with you was the best. Every day since your death has been hell. Life has really lost its meaning and I wonder if I can ever be the same person you fell in love with. Not a single moment has gone by that I have not thought about you. I miss you. There have been so many times I wanted to give up. There are still days when I feel like just ignoring everything and becoming a hermit out in the wild. I think the only reason I've kept on going was knowing how much trust you put in me when deciding to move here. I know that was not easy. It felt like I would be failing you had I decided to quit. I did not want there to be any chance of you being disappointed in me. I don't believe in the afterlife but I hope I made you proud. I will forever be loving you.

Achheta, I'm sorry I did not try harder to spend more time with you when you started college in the US. I will always regret that. The melancholy of not being able to be with family week in week out is something all immigrants have to accept, but moments like your loss makes me question whether deciding to move to the US for my studies nine years ago was even worth it. I promise myself I will never say maybe next time to anything ever again.

There are several people without whom this PhD would not have been possible. This includes my wonderful advisors Ray and Sammy for their great mentorship, my family for shaping me to becoming who I am today, Anita, Tim, Niroj, and Zach for never saying no when I came to you with random questions, and Nicole for all the help you provided me with in regards to the data. Ray and Sammy–you two especially deserve long glowing remarks for everything you have done but for just this once, I want to be a bit narcissistic and leave the rest of this section to acknowledge myself.

The last three years of my graduate school has really tested me in ways where I've questioned everything. It was really hard to be motivated to do anything and only I know what emotions I've grappled with, and continue to grapple with, day in day out. However, as I'm writing this, I want future me to be proud of myself and of this work. This work has been very important to me because I've witnessed both how a quality education, and a lack of it, can radically shape one's life. Life's unfair, and the goal was always to try to make the world a better and fairer place. I don't know what the future holds, but I hope I am reminded of this everytime I read this page.

Yamuna, Anu aunt, Roshan Dai, Ba – it's cruel how life gave you the short end of the stick, but the perseverance you showed gives me strength. You all continue to remind me why I chose to pursue this line of work. On this note, I'd also like to acknowledge everyone silently grieving while the world continues to move on. Nobody really sees the pain you hold so huge respect to everyone out there still carrying on during times when everything seems lost.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

What Happens When We Pay Our Teachers More? Evidence from New Jersey Public Schools

This paper examines the impact of increasing teacher salaries on student outcomes by exploiting variation from the "50K The First Day" campaign that established a $50K salary floor for new teachers across New Jersey school districts. Using school-level data from 2003 to 2019, I employ a staggered difference-in-differences (DiD) approach and first show that the campaign raised salaries for all teachers in New Jersey by approximately $1.5K. The results indicate that districts implementing the salary increase experienced improvements in 4^{th} grade and high school Math and English Language Arts (ELA) proficiency scores. I also observe modest gains in graduation rate and college enrollment. Analyzing the mechanisms through which these positive effects could have been observed, I rule out teacher migration as a key driver suggesting that the observed improvements are more likely due to changes in teacher motivation and the quality of new teachers entering the profession. Lifting teacher salaries for all teachers—regardless of their performance level—seems to be improving student outcomes in New Jersey.

1.1 Introduction

Schoolteachers in the United States are central to the education system, and their compensation continues to spark debate among policymakers and researchers. While prior studies have extensively examined factors such as teacher turnover, experience, and administrative leadership in shaping student outcomes,[1] the causal relationship between teacher salaries and student performance remains underexplored. The existing literature is characterized by mixed findings, leaving policymakers with little consensus on whether uniformly raising teacher salaries leads to measurable improvements in educational outcomes.

This paper seeks to address this gap by studying the impact of across-the-board teacher salary increases on student performance in New Jersey public schools. Specifically, I analyze the effects of the "50K The First Day" campaign, which mandated a $50K minimum salary for new teachers across all participating districts. By leveraging variation in the timing and magnitude of these salary changes, I employ a staggered Differences-in-Differences (DiD) under a continuous treatment framework to estimate the causal effects of salary increases. The results indicate that raising teacher salaries improves 4th grade and high school proficiency scores in Math and English Language Arts (ELA), along with modest gains in high school graduation rates and college enrollment.

This study contributes to the literature in three primary ways. First, I provide robust evidence on the impact of uniform salary increases, as opposed to performance-based pay schemes.[2] Unlike prior work that often focuses on selective pay increases targeting high-performing teachers, this paper investigates the effects of a universal salary floor applicable to all educators. The findings align with Han and Garcia (2022), who showed that higher

[1]See Eberts and Stone (1988), Miller (2013), Ladd and Sorensen (2017), Henry and Redding (2020), and Ng (2024) for relevant literature.

[2]See Fryer (2013), Goodman and Turner (2013), and Biasi (2021) for performance-based pay studies in the US, and Muralidharan and Sundararaman (2011), Muralidharan et al. (2016), and Hanushek et al. (2019) for studies outside the US context.

base salaries are associated with improved test scores, but I extend this work by employing a staggered DiD approach, offering stronger causal identification.

Second, I examine potential mechanisms driving these improvements, ruling out teacher migration as a key factor. The analysis suggests that the observed gains are more likely attributable to increased teacher motivation and higher-quality new hires. This contrasts with findings from Baron (2018) and Biasi (2021), who observed negative spillovers of performance-based pay systems in Wisconsin. By focusing on a union-led salary increase in New Jersey, this paper offers a complementary perspective, highlighting how collective bargaining agreements can improve outcomes without exacerbating inequalities between districts.[3]

Third, I conduct a cost-benefit analysis of salary increases by situating the findings within broader literature on school spending. Studies such as Jackson et al. (2016) and Jackson and Mackevicius (2024) document the positive effects of increased school budgets on student achievement, emphasizing reduced class sizes and better teacher retention as primary drivers. While the results support these findings, I further isolate the role of teacher salaries by showing that other school expenditures remained stable, ensuring that the observed effects are directly linked to the salary increase.

The remainder of this paper is structured as follows. Section 1.2 provides an overview of the "50K The First Day" campaign. Section 1.3 describes the data sources, followed by the empirical strategy in Section 1.4. Section 1.5 presents descriptive statistics, while Section 1.6 examines salary differentials between treated and untreated districts.[4] Section 1.7 evaluates district-level spending and school composition, and Section 1.8 discusses the effects on student outcomes. Section 1.9 compares the findings with other studies, and Section 1.10 explores mechanisms that may explain the observed effects.

[3] Appendix 3.6 briefly outlines the theoretical model exploring this mechanism.

[4] Treated districts are those that implemented the salary floor in a given year, while untreated districts had yet to do so.

1.2 Background: "50K the First Day Campaign"

In New Jersey, every school district uses a salary guide with steps that dictate teacher compensation. Table 1.1 shows an example salary guide where a teacher at Step 1 (no prior experience) with a Bachelor's degree will start at $45K annually. If they hold a Master's degree, they will start at $46K, and with a Master's plus an additional 30 credits, their starting salary will be $46.5K. While these guides serve as references, actual rates can vary across districts. The steps are not tied to specific subjects, though under rare circumstances, districts struggling to hire for in-demand subjects (e.g., Math) may offer higher step placements to attract candidates or offer them signing bonuses.[5]

Table 1.1: Example salary guide level.

Step	Bachelors	Masters	Masters + 30 Additional Credits
1	45,000	46,000	46,500
2	45,500	46,500	47,500
3	46,000	47,000	48,500

The paper exploits changes in the salary structure brought about by a bargaining negotiation led by the New Jersey Education Association (NJEA). NJEA is the leading labor union for NJ public school teachers, working to advance the rights, benefits, and interests of its members while promoting quality public education.[6] One of its notable campaigns was "$50K The First Day," which aimed to set a minimum salary of $50K for new teachers at Step 1 of the salary schedule.[7] Implementing such a guide would require all schools in a district to pay Step 1 teachers at least $50K. Figure 1.1 shows the share of school districts

[5]While not shown in the paper, on average, we do not find instances where a school offered a math/science teacher more than $2K above the standard step salary for the same experience level.

[6]After the U.S. Supreme Court's Janus v. AFSCME decision in 2018, teaching staff are no longer required to be members of the NJEA. NJEA continues to represent teachers across all districts, with little evidence suggesting that it advocates for improved working conditions only in specific districts.

[7]The NJEA is currently advocating for a "$60K The First Day" campaign, urging districts to raise the starting salary to $60K.

implementing the 50K salary in 2010 and 2015 respectively. Table A1 in Appendix shows in detail the years each school negotiated the settlement. Minimum salaries negotiated by different school districts varied from $50K to a little under $54K (Figure 1.2). I account for this variation in treatment intensity in the empirical model. From the figure, we see that most districts had a starting salary in the ranges of [50K, 52K] when they first negotiated the settlement.

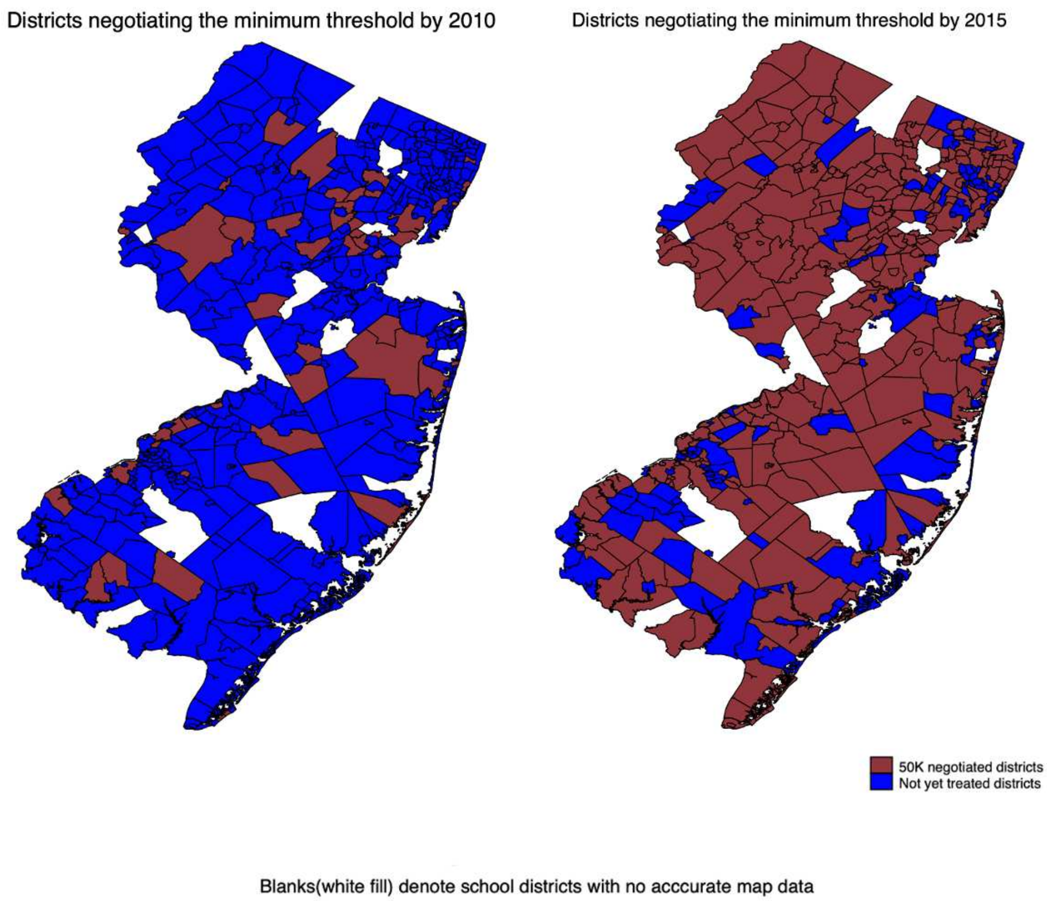


Figure 1.1: Number of districts passing the negotiation over the years.

Note: Figure shows the distribution of school districts that negotiated the 50K salary schedule in 2010 vs 2015. Negotiation always occurs at the school district level. School districts with a Step 1 Salary of at least 50K are colored brown. I exploit the staggered nature of this campaign to estimate the impact of salaries on student outcomes.

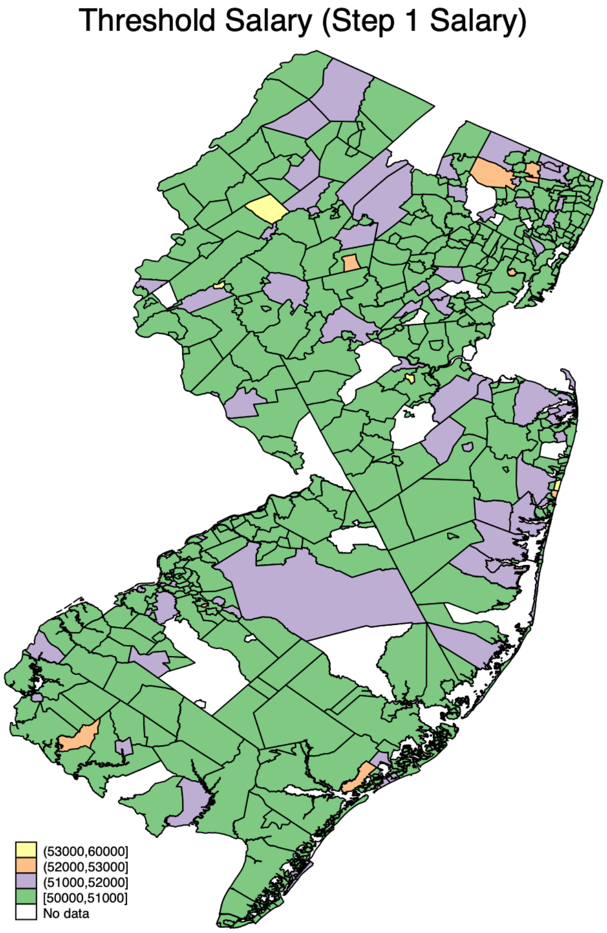


Figure 1.2: Step 1 Salary

Note: Figure shows the minimum Step 1 salary each school district had when they first negotiated this settlement. Negotiation always occurs at the school district level. I account for the difference in intensity of the treatment (salary offered) to estimate the impact of salaries on student outcomes.

New Jersey public schools implement salary guides at the school district level, raising the question of why some districts adopt new schedules while others do not. A common assumption is that districts refrain from increasing salaries due to budget constraints. However, NJEA argues that school districts often have the funds to reallocate toward salary increases but are deterred by political considerations. The decision to adjust salary guides is frequently driven by concerns over voter reactions, potential backlash from reallocating funds away from veteran teachers toward newer hires, or reluctance to approve settlement percentages above the average. Negotiations are particularly contentious when the policy is perceived as "taking money from the top," i.e., reducing veteran teacher pay.[8]

Rather than relying on expanded budgets, I show that many districts negotiated higher salaries by reallocating breakage — funds generated when a higher-paid retiring or resigning teacher is replaced by a lower-paid new hire. This strategy allowed districts to fund salary increases without requiring substantial new financial resources. My analysis in Section 1.7 indicates that treated districts hired fewer staff both before and after implementing the $50K minimum salary, suggesting that districts managed wage increases partly by controlling head-count rather than increasing overall expenditures. Additionally, the teacher-student ratio did not increase during this period, and declining student populations and the replacement of higher-paid, part-time staff facilitated the cost reallocation.

Although exploring the political dynamics that drove some districts to adopt the salary changes while others did not is beyond the scope of this paper, I acknowledge that broader state-level policies influenced New Jersey public schools during the study period (2003-2019). The first is the New Jersey School Funding Reform Act (SFRA), passed in 2008, which directed more state aid to districts unable to raise sufficient local revenue.[9] Another key policy was P.L. 2011 c. 78, signed by Governor Chris Christie, which increased public school staff pension contributions and significantly raised healthcare costs for teachers (New Jersey Department of the Treasury (2011)). While this policy may have reduced the appeal of teaching, I find only a small reduction in the number of teachers hired after its implementation, and thus do not control for it in the main empirical specification.

[8] I thank Crystal Inman of NJEA for providing background and context.

[9] In the robustness tests, I control for schools based on the socio-economic characteristics of the district they are situated in.

1.3 Data Description

For this study, I collect data from various sources to construct three primary datasets. First, I compile data from the New Jersey Department of Education (NJDOE) website to obtain school and school district level characteristics. This includes information on enrollment, demographics, average English Language Arts (ELA) and Math scores, graduation rates, and budget expenditures. Second, I use individual teacher-level data obtained from NJDOE through a formal data request, which contains detailed characteristics for all public school teachers in New Jersey. Finally, I use data provided by the New Jersey Education Association (NJEA) to identify when each school district adopted the $50K salary policy for new teachers. Combining all of these data, I construct the following datasets.

Individual Teacher-Level Dataset

This dataset contains comprehensive information on all public school teachers in New Jersey from 2004 to 2018, including variables such as teacher names, job categories/codes, experience levels, and highest degrees earned. Although individual teacher names are present, the dataset does not include a unique teacher tracking identifier, and name consistency is a significant issue, particularly in cases where last names change (e.g., due to marriage). Consequently, the dataset is inherently cross-sectional, limiting its capacity to track teachers across multiple years.

I primarily use this dataset to evaluate the impact of the salary negotiation campaign on teacher compensation. However, given the inconsistencies in naming conventions and the absence of a truly unique identifier, I also employ Natural Language Processing (NLP) techniques and Levenshtein distance to create a panel dataset of teachers. This approach enables me to analyze potential mechanisms by constructing a pseudo-panel that links teachers across years. I discuss this in more detail in Section 1.10.

District-Level Dataset

This dataset spans all school districts in NJ from 2003 to 2019. It combines publicly available data from the NJDOE to assess district-level spending patterns in different areas before and after the salary negotiations.

School-Level Dataset

The ideal student outcome data for the analysis would be individual teachers' classroom scores, which I could then use to create a value-added estimate of an individual teacher's impact. However, given the unavailability of these data, I rely on public data from NJDOE.[10]

This dataset includes data on all NJ schools from 2003 to 2019, such as student outcomes (measured as the percentage of students meeting proficiency or higher), racial demographics, enrollment, the percentage of students eligible for free/reduced lunch, total number of teachers, and other school-level characteristics. As the state examination underwent changes during this period, I use the percentage of students meeting proficiency or higher in state exams as the measure of student performance.[11] I define academic years by their Fall semesters. For example, I define the 2012–2013 year as 2012.

The primary outcomes are Math and ELA proficiency scores for Grades 4, 8, and high school, as well as graduation rates. To ensure consistency in the analysis, I group schools into four categories: those that exclusively serve elementary (including Grade 4), those that serve middle grades (including Grade 8), those that serve high school students, and those that serve both Grades 4 and 8.

1.4 Empirical Strategy

I employ a staggered Differences-in-Differences (DiD) model with a continuous treatment variable to assess the impact of salary increases across New Jersey school districts. The "$50K

[10] The request for teacher linked student performance data was rejected by the NJDOE.

[11] New Jersey Department of Education. History of New Jersey's Statewide Assessments. Retrieved from https://www.nj.gov/education/assessment/history.shtml.

The First Day" campaign set minimum salary thresholds for new teachers, with the exact magnitude of these thresholds varying across districts. Table A1 outlines the rollout of the campaign, detailing the years when districts negotiated the inclusion of the $50K minimum salary in their salary schedules. To capture the variation in salary levels and timing, I treat Step 1 salaries as a continuous variable, enabling us to estimate how different levels of salary increases influence student outcomes.

The empirical strategy relies on the staggered adoption of treatment across districts to identify causal effects. For the primary model, I use the estimator proposed by de Chaisemartin and D'Haultfœuille (2020). In a truly continuous treatment model, no two groups receive the same dose of treatment. In the case of minimum salaries, it is more the case that schools get a multi-valued treatment. The estimator we use is specifically designed for staggered treatment settings with a multi-valued treatment which can be extended to a truly continuous treatment as well. In this paper, I use the terms multivalued and continuous treatment interchangeably. The assumptions required to interpret treatment as continuous are discussed in detail in de Chaisemartin and D'Haultfœuille (2020). As robustness checks, I also apply the binary-treatment estimators developed by Callaway and Sant'Anna (2021) and Dube et al. (2023). While these methods do not accommodate continuous[12] treatments, their results are consistent and included in the Online Appendix.[13] I also run the regressions excluding never adopting school districts.[14] Below, I detail the structure of the empirical models at the teacher, district, and school levels.

1.4.1 Teacher-Level Model

At the teacher level, I employ a dynamic Difference-in-Differences (DiD) model to estimate the impact of Step 1 salary threshold implementation on overall teacher salary in-

[12]In this paper, I use the term continuous/multi-valued treatment interchangeably.

[13]The robustness checks using binary treatment methods are discussed in Appendix 3.6.

[14]Although not shown in the paper for brevity purposes, the results hold when excluding these groups.

creases. Given that my dataset is cross-sectional rather than a true panel, the event study is estimated using individual teachers as cross-sectional observations, with treatment effects effectively averaged by school and errors clustered at the district level since this is the level at which the treatment takes place.

$$Y_{it} = \lambda_t + \sum_{k \neq 0} \beta_k (t - E_g = k) + \mathbf{X}'_{it}\theta + \epsilon_{it} \tag{1.1}$$

- Y_{it}: Outcome variable (e.g., teacher salary) for teacher i in year t.
- λ_t: Year fixed effects, capturing state-wide shocks in each time period.
- β_k: Event-study coefficients representing the impact of 50K Step 1 Salary implementation relative to the reference period $k = 0.(t - E_g = k)$ is a binary indicator for being k years from the treatment starting.
- $\mathbf{X}_{it}$: Vector of controls which includes teacher experience and highest degree earned.
- ϵ_{it}: Idiosyncratic error term.

Since the dataset is cross-sectional rather than a panel, teacher fixed effects are excluded from the specification. By maintaining the data at the teacher level rather than collapsing to the school/district level, the model preserves the granularity necessary to accurately control for teacher-level characteristics such as experience and educational attainment. This approach aligns with de Chaisemartin and D'Haultfœuille's framework (see Appendix in De Chaisemartin and d'Haultfoeuille (2024)) , which emphasizes potential biases introduced when covariates are aggregated or handled inappropriately in a staggered DiD setup. Collapsing the data instead of estimating the event study this way would slightly change the estimate results (particularly the standard errors). The event study structure includes five pre-treatment periods and seven post-treatment periods, excluding the baseline period ($k = 0$).

1.4.2 District and School-Level Model

At the district and school level, I estimate the effect of salary increases using a Difference-in-Differences (DiD) framework with a continuous (multivalued) treatment. The treatment variable—Step 1 salary—is defined at the district level and takes on a finite number of ordered values.

Treatment is considered zero in all periods prior to policy implementation and jumps to a positive, district-specific level (Step 1 Salary) upon adoption. Once adopted, the treatment level remains fixed for each district throughout the post-treatment period.

The general structure of the event-study specification is:

$$Y_{it} = \alpha_i + \sum_{k \neq 0} \beta_k \cdot D_{it} \cdot 1[\text{EventTime}_{it} = k] + \lambda_t + \epsilon_{it} \tag{1.2}$$

where:

- Y_{it} is the outcome for school or district i in year t (e.g., test scores or graduation rates if outcome at school level, per-pupil spending if outcome at district level),
- D_{it} is the district-level Step 1 salary (treatment),
- $1[\text{EventTime}_{it} = k]$ is an indicator for being k years from the district's treatment adoption year,
- β_k captures the effect of treatment at event time k,
- α_i and λ_t are school/district and year fixed effects,
- ϵ_{it} is the error term.

Standard errors are clustered at the district level, which is the level of treatment assignment.

It is crucial to emphasize that although Equations (1.1) and (1.2) resembles a standard two-way fixed effects (TWFE) DiD regression often used in empirical economics, it is not the estimation strategy employed in this paper. Instead, I use a flexible, group-time-specific DiD estimator implemented via the `did_multiplegt_dyn` Stata package, following de Chaisemartin and D'Haultfœuille (2020). Unlike the TWFE OLS estimator—which pools all units across time and implicitly imposes homogeneous and linear treatment effects—this approach avoids forbidden comparisons (such as comparing already-treated units to later-treated ones) and constructs treatment effects using valid comparisons between treated and not-yet-treated districts at each calendar period (de Chaisemartin & D'Haultfœuille, 2020; Goodman-Bacon, 2021a). The traditional TWFE estimator is known to introduce negative weighting when adoption is staggered and treatment effects vary across groups or time. The method used here avoids such distortions.

The coefficients β_k and β_k are not directly estimated via OLS. Rather, they are constructed from multiple valid DiD comparisons across group-time cells and reflect average treatment effects on the treated at each event time and thus, they recover unbiased ATT effects. Recent contributions (Callaway et al., 2024; De Chaisemartin & d'Haultfoeuille, 2024; de Chaisemartin & D'Haultfœuille, 2020; Goodman-Bacon, 2021a) clarify how to recover unbiased estimates in this context by aggregating group-time-specific ATTs using appropriate weights.

Under heterogeneous treatment effects and multivalued treatments, β_k is better interpreted as a weighted average of ATTs across districts with different Step 1 salaries. In my empirical setup, I assume that once a district adopts a new salary schedule, the Step 1 salary remains constant in all post-treatment years. I impose this restriction for two key reasons. First, I lack longitudinal teacher-level salary records, so I cannot track salary trajectories or identify exact salary increases over time. Second, I do not observe pre-policy Step 1 salaries, necessitating the assumption that treatment is zero prior to adoption.

Even if richer data were available, estimating marginal or cumulative effects of evolving treatment intensity introduces substantial identification challenges. Endogeneity in treatment dosage paths remains a largely unresolved issue in the literature (Callaway et al., 2024; De Chaisemartin & d'Haultfoeuille, 2024). Nevertheless, the structure of my empirical framework is easily generalizable. Provided we have more granular data and can impose stronger assumptions about treatment evolution—future researchers can model the heterogeneous and cumulative effects of salary increases with greater precision and identify the per-dollar impact of salary changes on outcomes.

To assess the plausibility of identifying assumptions and visualize the dynamics of treatment effects, I present event-study plots of the estimated treatment effects. These plots track the evolution of outcomes relative to treatment adoption and allow for inspection of pre-trends. Before that, I begin with the descriptive statistics and summary trends in Section 1.5.

1.5 Summary Statistics

I now present descriptive statistics that illustrate different school-level characteristics, student composition, and educational outcomes. These summary statistics provide a baseline understanding of the sample used in the empirical analysis, helping to identify the key features of the schools included in the study and allowing us to observe trends and differences across years.

Table 1.2 displays the number of total observations for all schools by type. I distinguish between elementary, middle, and secondary schools, with an intermediate category for schools covering grades 4 to 8. Each school type is defined based on the grade levels it includes. Schools classified as "Four to Middle" serve students from Grade 4 in addition to the traditional middle school grades (grade 6 to 8).

Table 1.2: Number of Total observations from 2003-2019 by Grade Type Category

School Type	Frequency
Elementary	14,076
Four to Middle	4,437
Middle	4,947
Secondary	5,185

Note: The panel data used in the main results is strongly balanced, meaning that schools are consistently observed across all years. However, school outcomes are not always measured for all of these schools.

Table 1.3 shows the composition of students and average grade level scores by year. Over the sample period, I observe a steady decline in total enrollment and student-teacher ratios. The demographic breakdown reveals that the percentage of White students has steadily declined, while the percentage of Black students has remained relatively stable.

The table also summarizes key educational outcomes in this study. In this paper, I use percentage of students meeting proficiency level or higher as the measure for educational outcome. The data reveals notable variation over time, particularly post-2015, when New Jersey revised its core standards. This policy shift introduced consistently higher performance expectations, leading to a sharp decline in standardized test proficiency rates.

Given this drastic shift, I conduct robustness checks by restricting the analysis to pre-2015 data and standardizing scores by year. These adjustments ensure that the estimates are not solely driven by changes in educational standards. A potential concern is whether these shifts had heterogeneous effects across schools, disproportionately influencing some districts more than others.

To explore this, Figures A1 and A2 in the Appendix section depict the evolution of math and reading scores. The observed trends suggest that changes in scores—both declines and subsequent increases—follow consistent patterns across treated and not treated schools.

Table 1.3: Detailed Summary Statistics (School Level)

Year	Total Students	Student-Teacher Ratio	% White	% Black	Lunch Aid	Math 4	Math 8	Math HS	ELA 4	ELA 8	ELA HS	Grad.	Post Enroll.
2003	645.40	13.22	63.39	14.86	23.91	70.44	60.04	67.43	82.58	75.52	81.71	86.02	82.58
2004	649.44	13.13	62.45	14.96	23.89	74.21	64.58	71.47	85.79	73.08	83.72	85.94	83.46
2005	649.72	12.97	61.59	15.09	25.06	81.38	64.71	76.58	84.79	73.13	84.33	87.03	84.19
2006	648.00	12.84	60.64	15.15	24.52	83.42	66.41	77.05	83.34	75.01	84.86	89.09	84.40
2007	641.74	12.77	59.71	15.12	25.49	85.66	70.66	74.36	83.90	74.12	86.37	89.82	86.06
2008	635.82	12.61	58.88	15.04	26.16	85.63	68.99	75.86	85.37	82.04	84.11	89.58	85.58
2009	633.73	12.43	58.02	15.00	28.20	74.44	71.95	73.63	68.01	82.32	84.62	89.63	84.79
2010	631.92	12.43	57.34	14.79	29.79	78.43	68.99	74.96	64.38	82.61	88.18	90.70	86.09
2011	625.53	12.33	56.35	14.55	30.06	80.60	71.97	76.14	67.88	82.26	90.54	88.60	.
2012	622.32	12.77	55.58	14.35	31.58	79.14	73.00	83.03	64.43	82.61	91.72	88.65	.
2013	625.00	12.68	54.81	14.19	32.98	79.99	70.38	84.48	65.46	81.90	92.20	89.35	76.19
2014	621.14	12.43	54.13	14.01	34.03	76.53	71.76	83.64	65.64	79.67	91.79	89.88	76.68
2015	616.65	11.90	53.14	13.72	33.53	44.35	30.04	42.11	57.39	53.85	38.50	90.71	77.11
2016	614.48	12.02	52.19	13.52	33.38	49.87	31.55	44.52	59.46	56.28	44.92	91.22	77.73
2017	612.65	11.84	51.27	13.37	33.45	50.42	32.77	47.37	61.88	59.48	47.15	91.29	76.64
2018	609.51	11.78	50.26	13.16	32.94	52.08	33.58	47.18	63.44	60.89	51.38	91.73	78.69
2019	603.10	11.75	49.37	12.96	32.37	53.69	35.12	50.39	62.72	63.78	58.24	91.86	77.58
Total	628.60	12.46	56.42	14.34	29.51	70.77	59.53	67.59	70.96	72.98	75.79	89.48	81.18

Note: Table shows mean values at the school level. The fourth and fifth columns denote the racial composition of the student body, focusing on the percentage of White and Black students, respectively. The "Lunch Aid" column represents the percentage of students eligible for free or reduced-price lunch. Math and ELA denote the percentage of students meeting proficiency level or higher. "Grad." denotes high school graduation rates. "Post Enroll." refers to the percentage of students enrolling in post-secondary education within 16 months of graduation. In 2015, New Jersey's core standards were revised, impacting math and ELA test scores. As shown later in the paper, the results remain robust when restricting the analysis to data from years prior to 2015 and when standardizing the scores by year.

This consistency strengthens the argument that observed trends are not driven by systematic heterogeneity in test score changes.

I now need to test if early and late adopters are comparable since any systematic difference between early and late adopters could give us an estimate that is driven by select groups of school districts. The summary statistics in Table 1.4 provide essential context for understanding the comparability of districts treated at different time periods and highlight key differences with never-treated districts.

Table 1.4: Summary Statistics by Adoption Timing

Variable	**Early (pre-2010)**	**2011–2015**	**Late (2016+)**	**Never**
% Lunch Aid	31.20	29.17	25.34	67.88
	(30.63)	(27.91)	(21.79)	(18.71)
Student Enrollment Change	-2.15	-3.04	-2.73	0.46
	(77.74)	(53.15)	(44.92)	(52.70)
% Budget from State	36.4	31.7	28.4	67.6
	(28.4)	(25.0)	(16.9)	(16.2)
Total Students	667	636	572	736
	(435.0)	(434.7)	(387.6)	(525.8)
Total Per-Pupil Spending	13,376	13,134	12,166	13,379
	(2,950.6)	(3,114.3)	(2,767.0)	(2,822.3)
Classroom Per-Pupil Spending	7,373	7,423	6,896	7,394
	(1,435.1)	(1,674.5)	(1,489.5)	(1,453.8)
Starting Step 1 Salary ($)	50,528	50,503	50,459	—
	(497.6)	(511.9)	(572.3)	—
District Factor Group (1–8)	4.5	4.5	4.5	1.9
	(2.5)	(2.1)	(1.8)	(0.8)

Note: Standard errors are in parentheses. "Early" refers to districts adopting salary thresholds before or during the 2010-2011 period. "Late" refers to those adopting them in 2016-2017 or later. "Never" indicates districts that did not adopt the salary thresholds. Starting salary data for "Never" adopters is unavailable. "Lunch Aid" represents the percentage of students eligible for free/ reduced-price lunch. "Student Enrollment Change" reflects the average change (raw count) in student enrollment.

As demonstrated, early-treated (pre-2010), intermediate-treated (2011–2015), and late-treated (2016+) districts are similar in key observable characteristics, such as per-pupil spending, classroom-specific expenditures, and socioeconomic composition. For example, per-pupil spending varies modestly across treated districts, ranging from $12,166 in late-treated districts to $13,376 in early-treated districts, while classroom-specific expenditures per pupil also show comparable patterns. Furthermore, treated districts across all adoption periods have similar District Factor Group rankings[15] (averaging around 4.5), indicating comparable baseline socioeconomic conditions.

In contrast, never-treated districts are markedly different from treated districts. They rely significantly more on state funding (67.6% of their budgets) and have a much higher share of students eligible for reduced lunch which is a strong predictor of socio-economic status. These districts are disproportionately concentrated in the lowest socioeconomic categories, with an average District Factor Group of 1.9, indicating severe economic disadvantage compared to treated districts. These stark differences underscore the necessity of excluding never-treated districts from the identification of treatment effects, as their baseline characteristics deviate substantially from those of treated districts. Including them would likely violate the parallel trends assumption that is fundamental to the Difference-in-Differences (DiD) design. However, with only 39 never-treated schools (accounting for just 2.31% of the sample, as shown in Table A1), their exclusion is unlikely to materially affect the validity or generalizability of the estimates. This small proportion minimizes concerns about bias arising from their exclusion, ensuring the robustness of the results.

The comparability of treated districts across adoption periods reinforces the validity of the staggered DiD framework, as it suggests that the timing of treatment adoption is not systematically driven by pre-existing differences in district characteristics. Any remaining

[15]The District Factor Groups (DFGs) were first developed in 1975 for the purpose of comparing students' performance on statewide assessments across demographically similar school districts.

minor variations in spending or socioeconomic composition are effectively controlled for through district and year fixed effects in the empirical models.

I now turn to the analysis of how the "$50K The First Day" campaign impacted salaries across New Jersey school districts. In the next section, I estimate the causal effects of the policy on teacher and staff salaries and then estimate the impact of salary increases on student outcomes.

1.6 Effects of the Campaign on Salaries

In this section, I show the effects of the campaign on staff salaries. I begin by estimating Equation 1.1, restricting the sample to only Math and ELA teachers. The treatment here is binary.

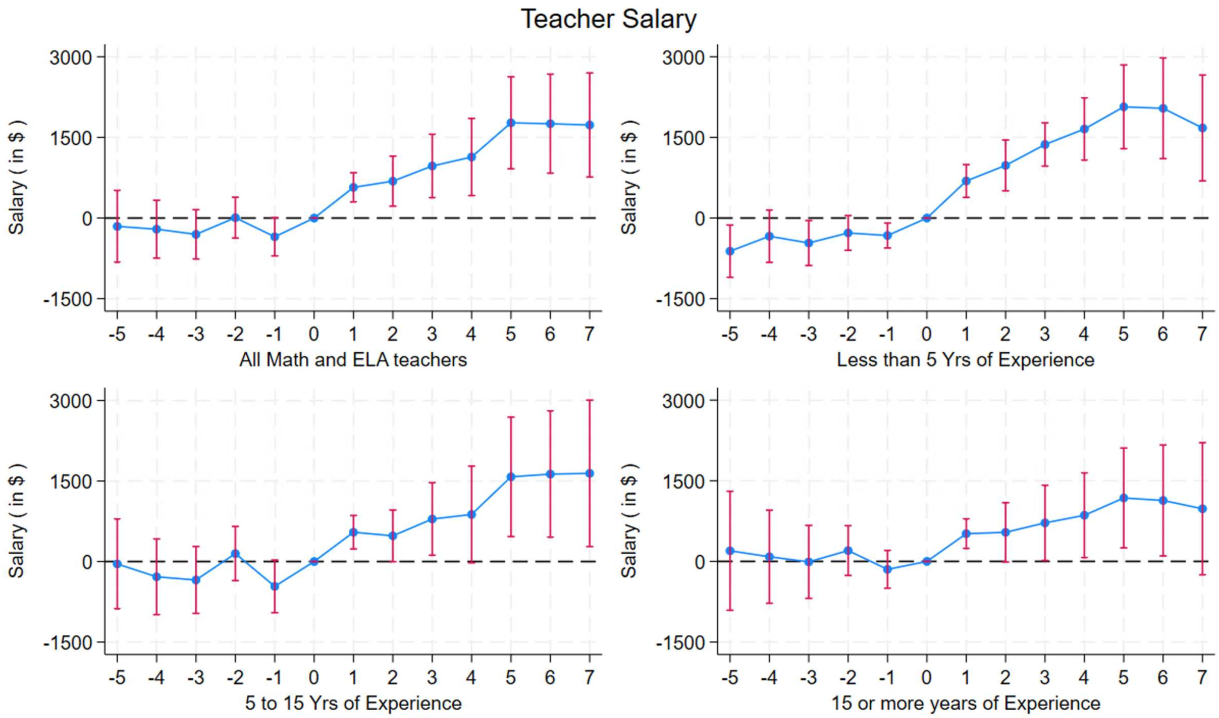


Figure 1.3: Salary of Math and ELA Teachers

Note: Point estimates and 95% confidence intervals of parameters β_k in Equation (1.1), constructed using valid Difference-in-Differences (DiD) comparisons between treated districts and not-yet-treated districts in the same calendar year. The event study shows how teacher salaries changed over time relative to the year the treatment (the salary campaign) was introduced. The x-axis represents the number of years before and after the salary increase took effect (with "1" being the year the policy was implemented), while the y-axis shows the estimated change in salaries. Each blue line represents the difference in salary for treated teachers compared to not-yet-treated teachers in that given year. The red bars show the 95% confidence intervals.

The first graph in Figure 1.3 (top left) presents an event study of all Math and ELA teachers. This event study shows how teacher salaries changed over time relative to the year the treatment (the salary campaign) was introduced, aggregating across all teachers regardless of experience level. The x-axis represents the number of years before and after the salary increase took effect (with "1" being the year the policy was implemented), while the y-axis shows the estimated change in salaries. Each blue line represents the difference in salary for treated teachers compared to not-yet-treated teachers in that given year. The red bars show the 95% confidence intervals.

Salaries begin to increase immediately after treatment (year 1) and continue to rise consistently in the following years. Before treatment (to the left of 1), the differences in salaries between treated and control groups were nearly identical—this supports the idea that treated and not-yet-treated schools were on parallel trends prior to the policy. I then break down the event study analysis based on the level of experience.

I now examine teachers with less than five years of experience (top right panel in Figure 1.3). These new teachers in treated schools see a significant salary increase relative to their peers in not-yet-treated schools. However, the parallel trends assumption could appear to be slightly violated for this group, as pre-treatment salaries show a small upward movement in treated schools before the salary floor was officially implemented. This could raise concerns about the validity of the results for this subgroup. To address these concerns, I apply the methods illustrated by Rambachan and Roth (2020) for robust inference in difference-in-differences and event study designs when the parallel trends assumption may be violated.[16]

When I break down the results by other experience groups (top right and bottom left panels), I observe positive salary increases for teachers with 5 to 15 years of experience and teachers with over 15 years of experience. For these groups, the pre-treatment trends align closely between treated and control schools, suggesting that the parallel trends assumption is satisfied. These teachers see salary increases averaging around \$1,000 to \$2,000, confirming that the salary campaign benefits teachers at all stages of their careers. This suggests that the policy successfully improved teacher salaries without creating potential discontent among senior staff members, who might have otherwise felt disadvantaged by a campaign that only benefited new hires. While breaking down the results by experience level (as seen in the

[16]Using this approach, I find that even when accounting for potential deviations in pre-treatment trends, the salary increases for new teachers are significantly higher than would be expected under linear extrapolation of pre-treatment trajectories. This provides strong evidence that the salary increases are indeed driven by the campaign and not by underlying trends.

other panels) provides valuable insights, the all-teachers analysis is particularly important because it allows us to assess the overall impact of salary increases across the entire group of teachers. This broader view helps us examine whether increasing salaries for all teachers could potentially improve student outcomes.

I next run the same analysis excluding Math and ELA teachers and including all other staff members (both teaching and non-teaching). The top panel in Figure 1.4 shows that other teaching staff members, such as science, history, or arts teachers, also experience similar positive impacts on their salaries after the campaign. I then restrict the sample to include only non-teaching staff (e.g., administrators, support staff). The event study (bottom panel) shows that these groups also experienced no negative changes in their salaries. Salary differences remained constant throughout the period of analysis. Taking stock, the event study results show that salaries increased for all subgroups, with new teachers experiencing salary increases the very year the negotiation was approved. Teachers with more experience saw increases the year after the act, and overall, no teachers were disadvantaged. Table A3 in the Appendix shows the average impact of the $50K The First Day Campaign on salaries for all full-time staff.

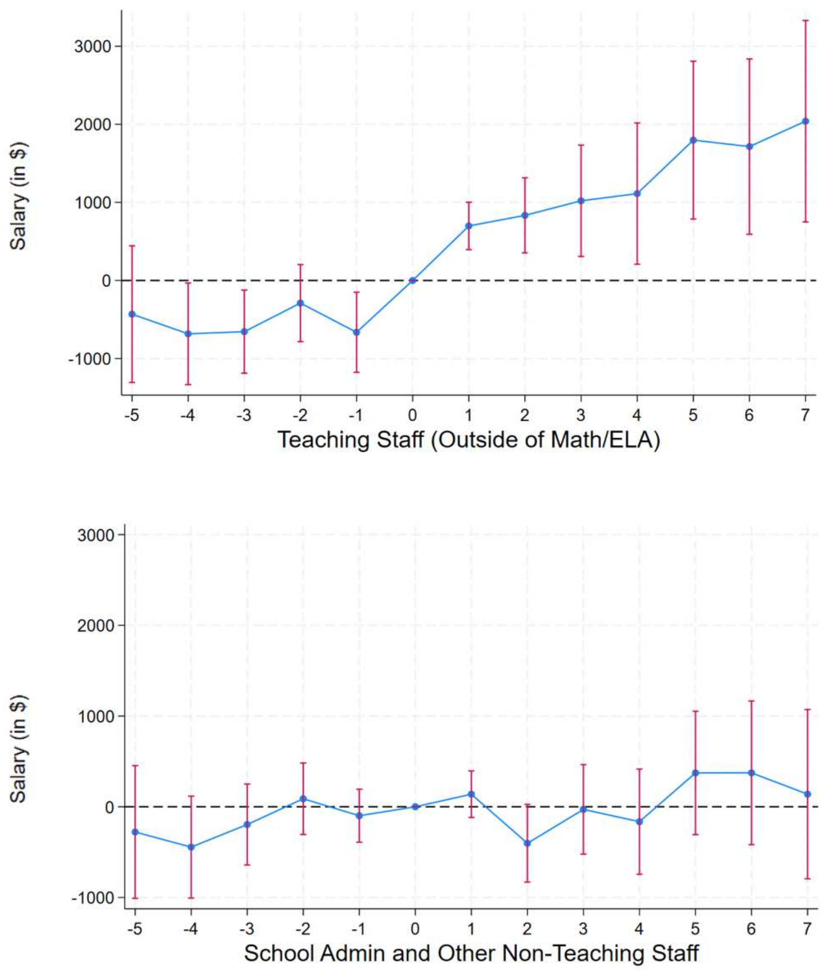


Figure 1.4: Salary of Teaching Staff Outside Math/ELA

Note: Point estimates and 95% confidence intervals of parameters β_k in Equation (1.1), constructed using valid Difference-in-Differences (DiD) comparisons between treated districts and not-yet-treated districts in the same calendar year. The y-axis represents the estimated changes in differences of yearly salary (measured in dollars) for treated versus not-yet-treated groups relative to the reference year (time period 0). The x-axis shows the time period relative to the treatment year, with 1 indicating the year the treatment took place.

Taken together, these findings allow us to rule out any negative spillover effects among different types of staff due to the salary increases for teachers. The absence of salary reductions or stagnation for other staff members, both teaching and non-teaching, suggests that the policy was implemented in a way that avoided internal disparities within the district's workforce. This minimizes the risk of unintended consequences, such as dissatisfaction among non-teaching staff or other educators not directly targeted by the salary floor. Given that non-teaching staff and administrators did not receive higher salaries due to this campaign, one potential issue could be that there is a negative effect on students due to their demotivation. If so, I argue that this would cause the estimated effects to be downward biased, as it is unlikely that these groups of staff would experience any changes that would positively affect student outcomes.

1.7 Effects on District Spending and School Composition: Where is the Money Coming From?

In this section, I examine the impact of the salary campaign on district-level spending and compositional changes at the school level. I begin by estimating Equation 1.2 at the district level to analyze spending patterns within a district before and after the campaign.

Figure 1.5: Difference in Per Pupil Spending (in Dollars at District level)

Note: Point estimates and 95% confidence intervals of parameters β_k in Equation (1.2), constructed using valid Difference-in-Differences (DiD) comparisons between treated districts and not-yet-treated districts. Unit of analysis is at the school district level. The y-axis shows changes in differences in yearly spending per pupil, and the x-axis refers to the time period before and after the treatment. The reference year is time period 0, which indicates the year before the treatment took place. Treatment is non-binary (continuous). The figure on the top left shows per-pupil spending in all areas. The figure on the top right shows per-pupil spending used only on teacher salaries. The bottom figure shows per-pupil spending outside of expenditures on classroom salaries.

Figure 1.5 shows the event study for per-pupil spending at the district level. The top-left panel demonstrates that total per-pupil spending did not significantly increase in treated schools, suggesting that districts did not receive or allocate additional overall budget resources as a result of the treatment. The bottom panel further indicates that differences in spending on non-salary components remained largely unchanged, implying that changes in spending outside of teacher salaries could not have driven the observed outcomes. The top-right panel suggests that differences in per-pupil spending on teacher salaries did not increase post-treatment. This may initially appear counterintuitive given that teacher salaries increased. However, this result can be explained by several factors.

Firstly, the difference in the total amount spent per student on teacher salaries between treated and not-yet-treated schools did not change significantly due to shifts in staffing patterns. Specifically, the number of teaching staff (total teachers) hired before districts implemented the salary schedule was on a decreasing trend in treated districts, as shown in Figure 1.6. This compositional change, where schools hired fewer teachers or replaced departing, higher-paid senior staff with lower-paid new hires before implementing the new Step 1 salary, helps explain why total spending per student on salaries remained relatively flat, even though individual teacher salaries increased. This phenomenon arises from the fact that while individual salaries rose, the number of teachers in treated districts decreased, balancing out the total spending per student. Post treatment, we do not observe the same downward trend.

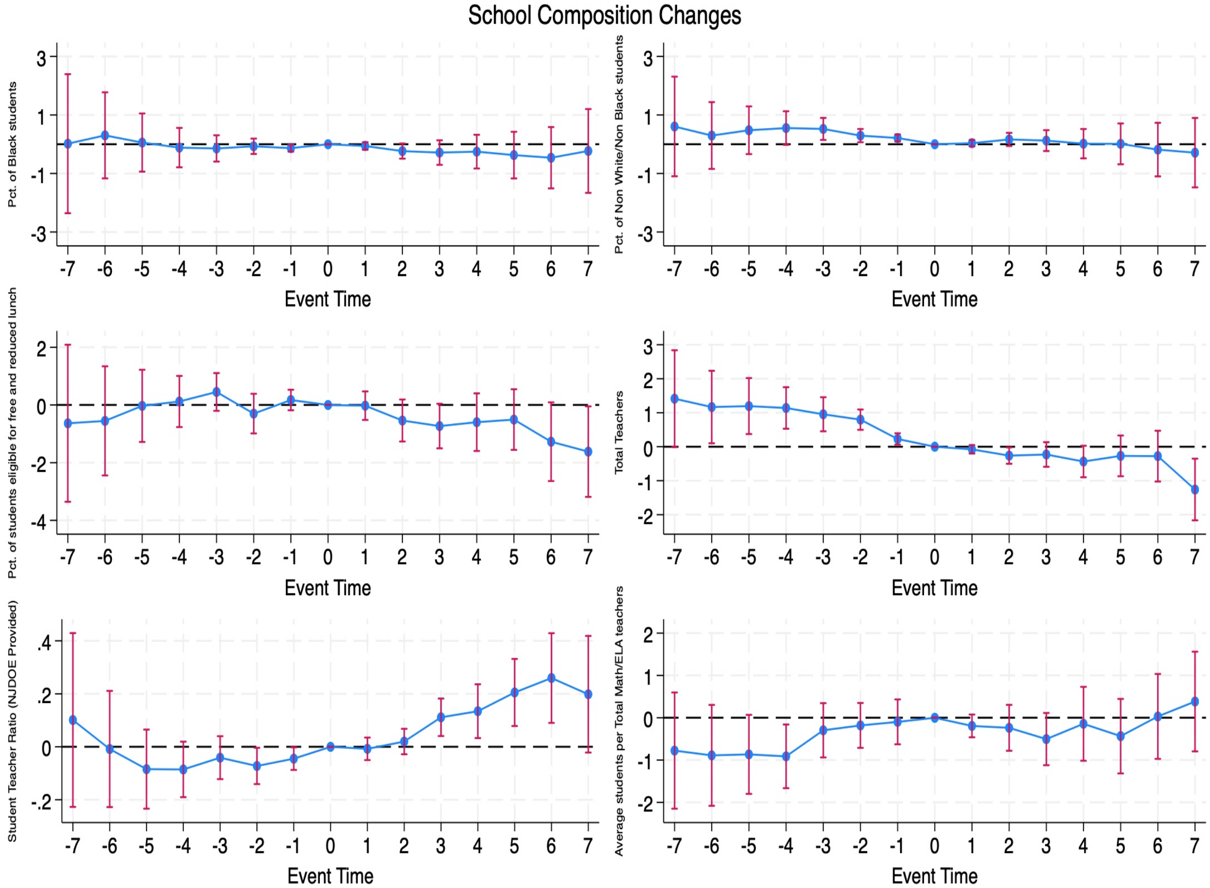


Figure 1.6: School-Level Changes

Note: Point estimates and 95% confidence intervals of parameters β_k in Equation (1.2), constructed using valid Difference-in-Differences (DiD) comparisons between treated districts and not-yet-treated districts. Unit of analysis is at the school level. The y-axis shows yearly outcomes, and the x-axis refers to the time period before and after the treatment. The reference year is time period 0, which indicates the year before the treatment took place. Treatment is non-binary. The racial composition of students remains fairly similar, and the total number of staff members employed (bottom right) decreases.

Figure 1.6 also highlights the compositional changes in student demographics at the school level. The racial composition of students (top panels) remains consistent before and after the campaign, suggesting no significant demographic shifts that could confound the analysis. The percentage of students eligible for free and reduced lunch remains fairly stable, with only about a 1% drop in treated schools. Before treatment, treated schools had more teachers than not-yet-treated schools, which explains why total spending per student on salaries remained relatively flat. Hiring fewer teachers could increase the burden on teachers in treated schools, and I do observe a slight increase in the student-teacher ratio in treated schools post-treatment (bottom left panel). This indicates that even though total enrollment in schools was decreasing (Table 1.3), the student-teacher ratio was higher in treated schools. Although this increase is minimal (less than one additional student per teacher), an increase in the student-teacher ratio could potentially affect student outcomes.

If this compositional change were to influence student outcomes, I argue that it would likely introduce a downward bias. Following the findings from Angrist and Lavy (1999), a higher student-teacher ratio generally leads to lower student outcomes, making any potential improvements in student performance even more noteworthy given this slight increase. However, newer studies, including one by Angrist himself, suggest that class size has no meaningful effect on student outcomes (Angrist et al., 2019; Leuven & Løkken, 2020).

Importantly, the ratio of students to Math and ELA teachers—the subjects analyzed in this study—remains stable throughout the study period.[17] This suggests that Math and ELA teachers in treated schools were not overburdened with more students after the campaign.[18] There is a small possibility that the increase in the student-teacher ratio might have negative spillover effects on other subjects (mainly those outside of Math, ELA, Science, and Social Studies). However, given that 1) the increase in the student-teacher ratio is very small (less than 0.5 students per teacher) and 2) recent studies show no negative effects when class sizes are reduced by up to three students, it is unlikely that our estimates are biased due to the teacher-student ratio. Nevertheless, given the absence of outcome data for other subjects, we cannot entirely rule out such effects.

[17] While not shown in the paper, the ratio of students to Science and Social Studies teachers also remains stable. The overall increase in the student-teacher ratio in treated schools is driven by changes in other subject areas.

[18] Although not shown for brevity, I conducted the event study on other areas of spending and found no meaningful changes due to this campaign.

Overall, the results presented here address several key concerns about resource reallocation and staffing composition changes that could potentially confound the later analysis of student outcomes. Specifically, while individual teacher salaries increased, the total per-pupil spending on salaries remained stable due to reductions in staffing levels, effectively mitigating the risk of budgetary distortions. Additionally, the stability in the ratio of students to Math and ELA teachers suggests that core subject teachers were not overburdened, minimizing the risk of increased workloads that could negatively impact student performance.

Moreover, potential endogeneity issues—such as schools adjusting their spending patterns in other areas or experiencing demographic shifts—are largely ruled out. The compositional stability of the student body and the lack of significant changes in district spending on non-teaching resources ensure that the analysis remains focused on the causal effect of salary increases rather than confounding factors. These findings lend credibility to the robustness of the identification strategy, allowing us to more confidently assess the impact of teacher salary increases on student outcomes. With these considerations in place, I now move on to examine whether higher salaries lead to improvements in student performance.

1.8 Main Results - Impact of the Campaign on Student Outcomes

This section presents the main results of the paper where I estimate Equation 1.2 using individual schools as the unit of analysis. The key outcomes of interest are Math and ELA proficiency levels across grades 4, 8, and high school, as well as graduation rates and college enrollment.

1.8.1 Math Outcomes

I begin by examining the impact of the salary increases on Math proficiency outcomes, with the results illustrated in Figure 1.7. This figure presents an event study of Math proficiency levels for 4th grade, 8th grade, and high school, providing a detailed look at how these outcomes evolved before and after the salary increases were implemented.

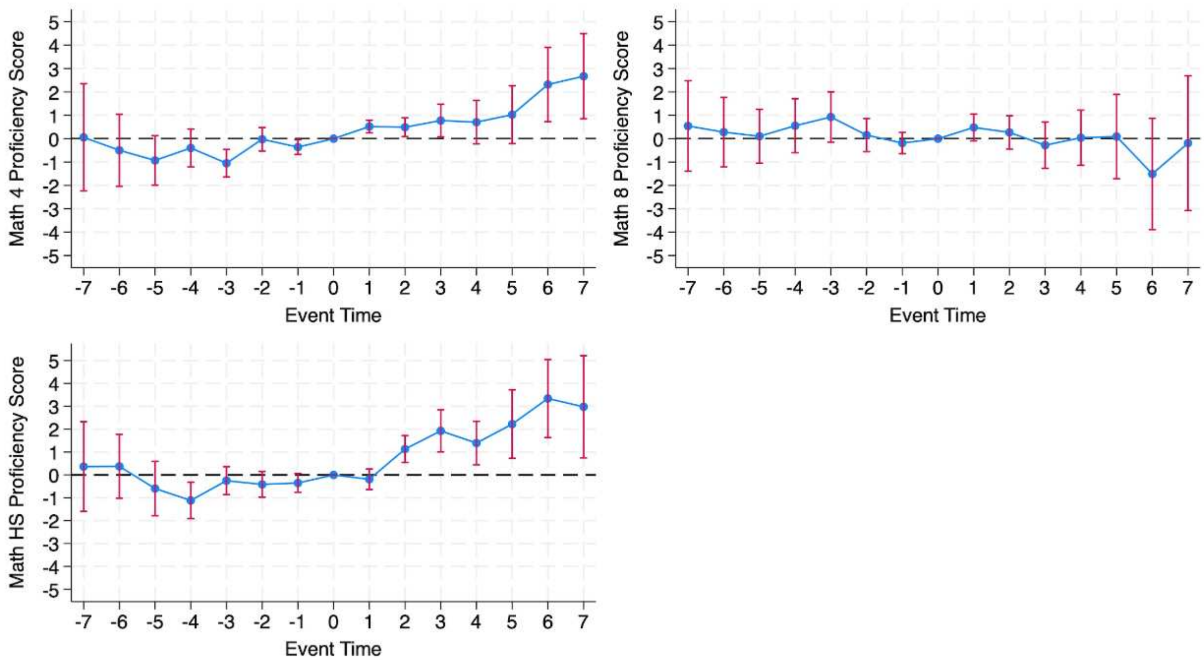


Figure 1.7: Proficiency Level Met (Math)

Note: Point estimates and 95% confidence intervals of parameters β_k in Equation 1.2 at the school level, constructed using valid Difference-in-Differences (DiD) comparisons between treated districts and not-yet-treated districts. The y-axis shows changes in proficiency level (expressed in percentage points), and the x-axis refers to the time period before and after the treatment. The reference year is time period 0, which indicates the year before the salary campaign took place. Treatment is non-binary.

In interpreting the event study estimates, the horizontal axis represents time relative to the introduction of the salary increases (with year "0" being the treatment year), and the vertical axis represents changes in the percentage of students meeting Math proficiency standards. The blue dots in the figure show the estimated treatment effect at each time period, while the red bars represent the 95% confidence intervals around these estimates. A value of 0 on the y-axis would indicate no change in differences in proficiency levels, while positive values suggest improvements in Math performance following the salary increases.

The pre-treatment period (to the left of year 0) shows that the estimated treatment effects are close to zero, which validates the parallel trends assumption—i.e., the proficiency levels in treated and control schools were evolving similarly before the policy was implemented. This is crucial for establishing that the post-treatment effects can be attributed to the salary increases rather than to pre-existing differences between treated and not-yet-treated schools.

After the salary campaign, I observe a significant and sustained increase in Math proficiency, particularly in 4th grade and high school. The pattern for high school Math is consistent with that of 4th grade, demonstrating that the salary increases benefited students across multiple educational levels. Importantly, the effects appear to persist over time, with continued improvements in proficiency levels even several years after the salary increases were implemented. I do not observe a significant increase in 8th-grade scores post-treatment, and the estimates for 8th grade exhibit larger standard errors. These findings highlight the differential effects of the salary campaign across different grade levels, with the most pronounced improvements observed in early and late stages of schooling.

1.8.2 ELA Outcomes

I next turn to the event study results for ELA outcomes, as shown in Figure 1.8.

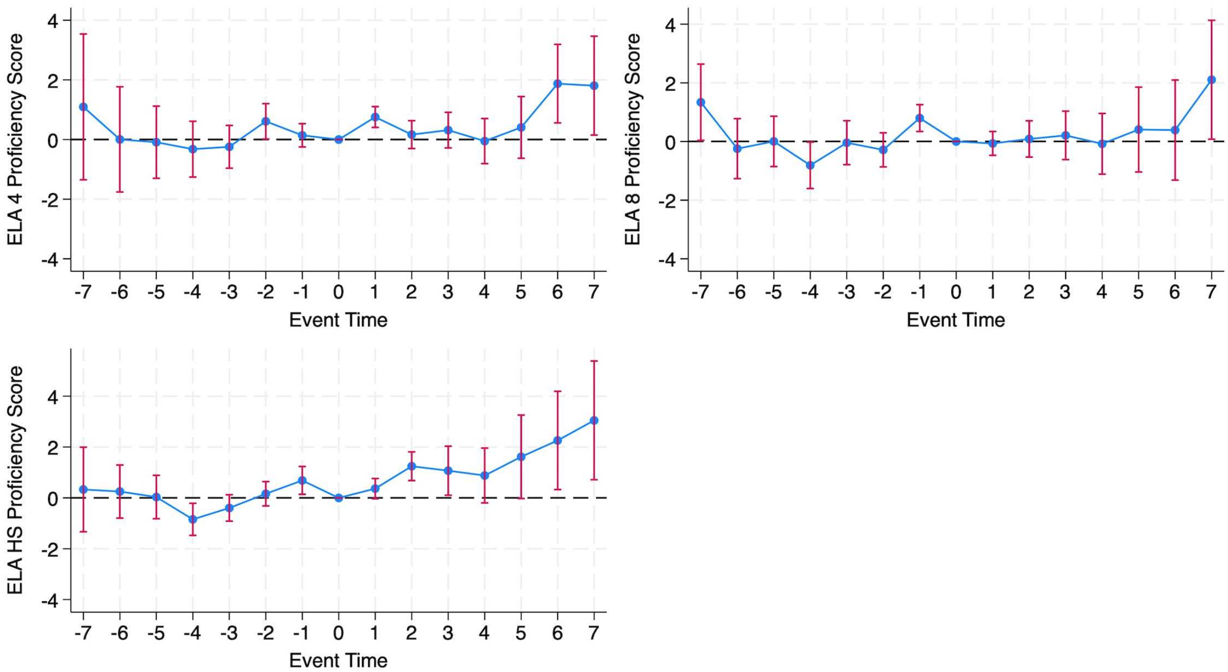


Figure 1.8: Proficiency Level Met (ELA)

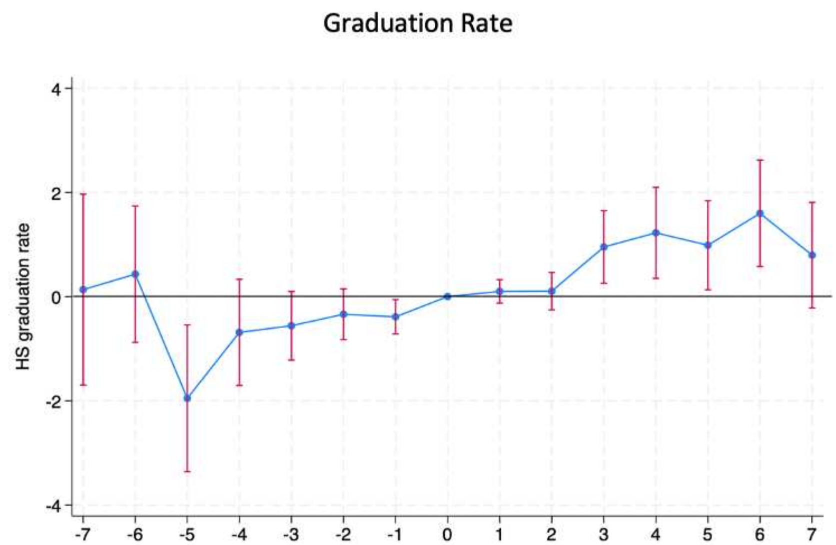
Note: Point estimates and 95% confidence intervals of parameters β_k in Equation 1.2 at the school level, constructed using valid Difference-in-Differences (DiD) comparisons between treated districts and not-yet-treated districts. The y-axis shows changes in proficiency level (expressed in percentage points), and the x-axis refers to the time period before and after the treatment. The reference year is time period 0, which indicates the year before the treatment took place. Treatment is non-binary.

The vertical axis shows changes in the percentage of students meeting ELA proficiency standards, where positive values reflect improvements in performance. The pre-treatment period (left of year 0) shows that ELA proficiency levels were evolving similarly in treated and not-yet-treated schools before the salary increases were introduced.

Following the implementation of the salary increases, the event study reveals a significant and sustained positive effect on ELA proficiency in both 4th grade and high school. The timing of the improvement in high school ELA is similar to that of 4th grade, with proficiency rates rising immediately after the salary campaign and continuing to increase over subsequent years. These results suggest that the policy's benefits extended across different educational levels, with students at both ends of the schooling spectrum benefiting from the enhanced teacher compensation. However, as with the Math outcomes, the event study for 8th-grade ELA shows no significant impact. The estimates for 8th-grade ELA proficiency exhibit larger standard errors, making it difficult to detect a clear pattern.

1.8.3 Graduation and College Enrollment

Figure 1.9 examines the effects of salary increases on high school graduation rates and college enrollment. This analysis allows us to assess the potential long-term educational attainment impacts beyond proficiency scores.



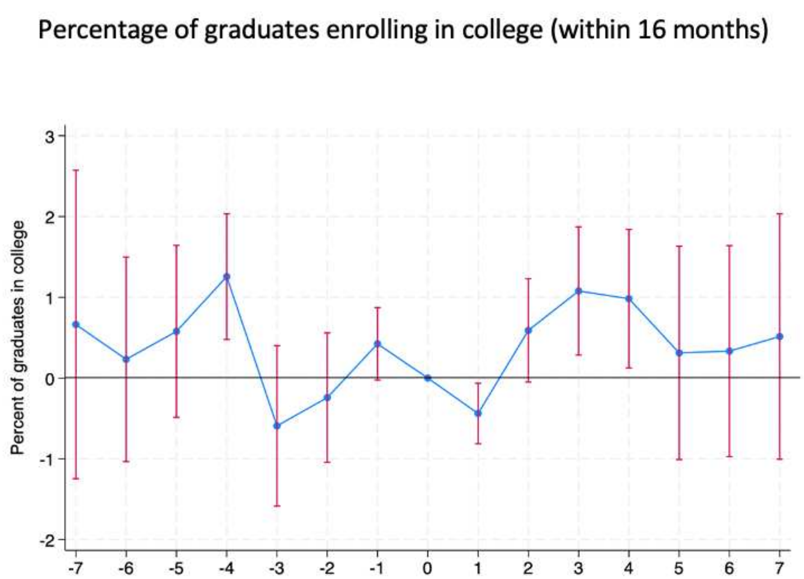


Figure 1.9: Graduation and College Enrollment

Note: Point estimates and 95% confidence intervals of parameters β_k in Equation 1.2 at the school level, constructed using valid Difference-in-Differences (DiD) comparisons between treated districts and not-yet-treated districts. The y-axis shows changes in graduation rates and college enrollment, respectively (expressed in percentage points). The x-axis refers to the time period before and after the treatment. The reference year is time period 0, which indicates the year before the treatment took place. Treatment is non-binary.

The event study in Figure 1.9 shows a positive and significant effect of salary increases on high school graduation rates. However, I do not observe a comparable upward trend in the percentage of graduates enrolling in college. While higher salaries may have improved school conditions and student performance, these improvements do not appear to directly translate into increased college enrollment. I detect some evidence of possible pre-trends in graduation rates, raising concerns about baseline differences before treatment. Furthermore, the lack of a consistently increasing pattern for college enrollment suggests that the effects may be more complex or influenced by other unobserved factors. For this reason, the primary focus remains on Math and ELA scores. Another limitation of focusing on graduation and college enrollment is the absence of post-secondary outcome data. While I can track how many students graduate and enroll in college, I lack information on their subsequent performance, degree completion, or labor market outcomes, which are arguably more meaningful indicators of long-term success. Without these data, it is difficult to fully assess the broader implications of salary increases on students' post-secondary trajectories.

1.8.4 Taking Stock

In this section, I summarize the key findings by quantifying the overall impact of the salary increases across student outcomes. Figure 1.10 (and Table A4) provides the average treatment effects over the seven-year period following the implementation of the salary campaign.

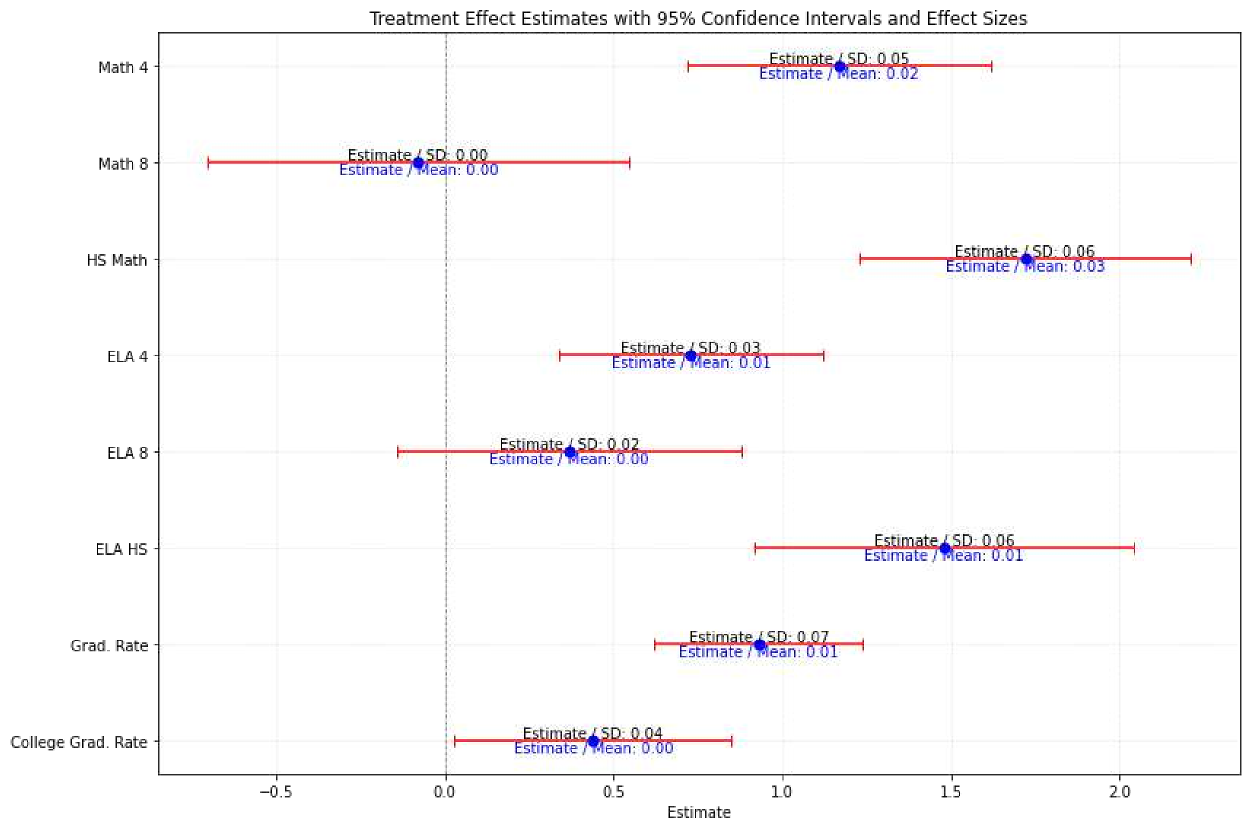


Figure 1.10: Estimation of Treatment Effects: Average Total Effect

---

Note: The figure presents the estimated treatment effects along with 95% confidence intervals for key student outcomes, including Math and ELA scores at different grade levels, high school graduation rates, and post-secondary enrollment rates. The X-axis shows the estimated effect size. The inclusion of 'Estimate/SD' and 'Estimate/Mean' provides a standardized metric for interpreting effect sizes, allowing for direct comparisons across outcomes with different scales. Value of 0.00 indicates effect is less than 0.01. Below, we present the mean and standard deviation for each outcome variable, enabling readers to verify the calculated effect sizes:

	Math 4	**Math 8**	**HS Math**	**ELA 4**	**ELA 8**	**ELA HS**	**Grad. Rate**	**College Enroll.**
Mean	70.77	59.53	67.59	70.96	72.98	75.79	93.83	81.18
SD	21.69	24.44	26.01	18.79	20.17	22.72	7.64	12.02

For instance, the 'Estimate/SD' for Math 4 is calculated as $\frac{1.17}{21.69} \approx 0.05$, and 'Estimate/Mean' is calculated as $\frac{1.17}{70.77} \approx 0.017$. Similar calculations apply to all other outcomes, allowing for straightforward verification and interpretation of effect sizes.

The estimates are presented in terms of standard deviations (SD) and average scores. I briefly discuss the magnitude of the effects by analyzing the estimates.

For 4th-grade Math, scores increase by approximately 0.055 SD on average, reflecting a substantial improvement in the proportion of students meeting proficiency standards. High school Math proficiency shows a similar positive trend, with scores increasing by about 0.06 SD post-treatment. In terms of the ratio to average scores, math scores for 4th grade and high school increased by roughly 0.02.

The results for ELA proficiency are also positive, though slightly smaller in magnitude compared to Math. In 4th-grade ELA, the average increase is around 0.045 SD, while high school ELA proficiency improves by approximately 0.06 SD. In terms of the ratio to average scores, ELA scores for 4th grade and high school increased by roughly 0.01.

Regarding graduation rates, I observe a meaningful increase of 0.06 SD post-treatment. This suggests that the salary increases not only impacted student performance in standardized tests but also had a lasting effect on high school completion, a critical indicator of long-term educational success. The impact on college enrollment, while positive, is less conclusive. I estimate an effect of around 0.04 SD, but the estimates become noisier in the years following the salary increases.

Overall, the evidence suggests that increasing teacher salaries led to meaningful improvements in student outcomes, with the most pronounced effects seen in 4th grade and high school, while 8th-grade outcomes remained largely unchanged. The magnitude of these effects suggests that improving teacher salaries could be a potentially powerful tool for boosting educational performance across different stages of schooling.

1.8.5 Robustness Checks, Heterogeneous Effects and Threats to Validity

To ensure the robustness of the findings, I apply several alternative specifications to verify the consistency of the results.

First, I restrict the sample to a strongly balanced panel, focusing only on schools that have complete data for all event-time periods. In line with the DiD literature, I refer to this approach as restricting to "same switchers." By including only those schools with data for all specified event-time periods, I avoid the potential for compositional bias that could arise if different schools contribute to different periods. This approach ensures that the treatment effects are estimated based on a consistent set of schools over time, which strengthens the internal validity of the results. However, this restriction reduces the sample size and could introduce selection bias, as the final sample may no longer be representative of the broader population of schools. Figure A4 shows that the results remain consistent under this more restrictive specification.

Next, I standardize the outcome variables to account for changes in testing standards that were introduced in 2015. This standardization ensures that any observed effects are not confounded by shifts in assessment criteria, allowing for a more accurate comparison of student performance over time. Figure A5 in Appendix illustrates the event study estimates using standardized outcome values.

Lastly, I employ two alternative methods for estimating the event study effects. The first, proposed by Callaway and Sant'Anna (2021), uses a group-time average treatment effect approach under a binary treatment setting. The second, introduced by Dube et al. (2023), uses local projections to estimate the treatment effects, providing a different way to handle staggered adoption by projecting the effects over time. Figure 3.6 in the Appendix section shows the event study of the outcomes under these approaches. The results remain robust across all specifications.

To explore potential heterogeneity in the effects, I estimate the model separately for schools grouped by the socio-economic status (SES) of their districts. This approach enables us to assess whether the impact of salary increases differs across schools with varying SES backgrounds. However, the estimates obtained from this disaggregated analysis are too noisy to draw definitive conclusions about differential effects based on socio-economic status since schools are divided into seven distinct groups.

I do control for SES in the robustness test, by using broader groupings of schools according to their relative socio-economic status, categorized based on the district factor group of the school districts. Under this broader classification, where schools are matched with other schools in the same socio-economic group, the results remain consistent, and in many cases, the estimates are more precise (see Figure A7). Furthermore, in all specifications where school-level controls are introduced—such as racial composition, student enrollment, and the percentage of the budget funded by the state—the estimated effects are generally stronger and more precise.

I also run the estimates by randomizing the year treatment took place by entering fake years of treatment. Figure A8 shows the event study for student outcomes where the treatment year is set to be three years later than the actual treatment year, and the treatment intensity is randomized between 50K to 60K. Under this setting, I find no significant effects on student outcomes further strengthening the validity of the estimates.

While the results indicate strong positive effects of salary increases on student performance, it is important to examine whether these effects are disproportionately driven by a specific subset of schools—particularly those that adopted the policy very early or late. If treatment effects are concentrated in a particular cohort, this could raise concerns about the broader applicability of the findings. To assess the robustness of the estimated effects across different groups, I also disaggregate the treatment effects by the year in which schools received the salary increase. This approach allows us to determine whether the observed improvements in student outcomes are consistent across all cohorts or whether they are influenced by specific groups of schools that implemented the policy at different times.

Figure A3 presents the estimated treatment effects for high school and 4th-grade Math and ELA scores across various treatment groups, alongside the overall average effect (GAverage). Notably, there is no clear temporal pattern in the estimated effects. If early adopters had consistently stronger or weaker effects relative to other groups, it would suggest that pre-existing differences may be driving the results. However, the relatively stable estimates across different treatment cohorts suggests that the effects are not disproportionately influenced by a single group of schools/ school districts. Instead, the results suggest that the positive impacts of salary increases on student performance are robust across different adoption periods, strengthening the overall validity of the findings.

While the empirical strategy accounts for a range of potential confounders, certain untestable assumptions remain, which, if violated, could pose threats to identification. If districts implementing larger salary increases differ in ways not fully captured by observable characteristics, the estimates may reflect both these underlying differences and the causal impact of salary increases. Additionally, while I assume a smooth relationship between treatment intensity and outcomes, a highly nonlinear or discontinuous response function could introduce bias if standard modeling choices fail to capture the true treatment effect. Finally, as with any staggered adoption setting, concurrent district– or state-level policy changes could confound the estimates if they differentially affect districts based on treatment intensity. The interpretation relies on the assumption that unmeasured shocks—such as statewide policy changes or economic fluctuations—do not differentially impact treated and untreated districts. Although I demonstrate that per-pupil spending and key demographic trends remain stable before and after treatment, I cannot rule out all possible confounding factors.

1.9 Comparison with Other Studies

In this section, I compare the magnitude of the estimates with findings from other relevant literature. The study focuses on a constant increase in salaries for all public school teachers in New Jersey. It is essential for readers to note that this approach differs from studies that evaluate performance-based salary schedules. Since this campaign did not distinguish between teachers, salary increases occur for all teachers, irrespective of their performance.

In two 2013 studies, Fryer (2013) and Goodman and Turner (2013) found that bonuses of $1,500 to $3,000 per teacher in New York City public schools had little impact on teacher effort, student performance in math and English, or classroom activities. In another study from Tennessee, Springer et al. (2012) found that students whose teachers were eligible to receive bonus payments performed at the same level as those whose teachers were ineligible, indicating no significant effect. Similarly, Biasi et al. (2021) noted that after Act 10 in Wisconsin, wage growth remained small and negative for teachers who stayed in their positions, but increased significantly for those who moved to new districts ($1,750 at the median). This led to flexible pay (FP) districts increasing reading and math scores by 0.04 and 0.06 standard deviations, respectively. However, salary changes varied in this study, and thus a direct comparison with the results may not be appropriate.

Baron (2018) examined the impact of Act 10 on average student achievement in Wisconsin, finding that the reduction in union power decreased teacher salaries by roughly 4% and reduced average Wisconsin Knowledge and Concepts Examination (WKCE) scores by approximately 20% of a standard deviation. I begin by comparing the magnitude of the estimates with these two Wisconsin-focused studies. Before delving into this comparison, I present the summary statistics of teacher salaries by year in Table 1.5.

Table 1.5: Summary of Salary by Year with Median, SD, and Number of Observations

Year	Mean Salary	Median Salary	Std. Dev.	N
2004	60,274	54,478	19,198	124,898
2005	61,882	55,655	19,471	126,748
2006	63,571	57,248	19,684	123,361
2007	65,264	58,936	19,705	125,249
2008	67,082	60,963	19,760	125,978
2009	68,992	63,239	19,735	126,319
2010	70,422	65,211	19,569	120,526
2011	70,823	65,643	19,383	120,725
2012	71,186	65,851	19,580	120,594
2013	72,077	66,998	19,596	123,871
2014	72,456	67,451	19,527	124,597
2015	73,051	68,200	19,503	123,485
2016	74,237	69,743	19,544	124,902
2017	75,436	71,375	19,592	123,971
2018	76,397	72,520	19,612	125,159
Total	69,523	65,188	20,141	1,860,383

Note: Table reports the average and median salary for all full-time staff in NJ public schools, rounded to integer values. Individual teacher-level salary data are winsorized at the 1st and 99th percentiles. The fifth column reports the total number of full-time staff (both teaching and non-teaching) in NJ public schools.

From Table A3, which shows the event study estimates of the campaign on all staff in a school, I observe that differences in salaries between teachers in treated and control schools increased by approximately \$1,200 on average. This corresponds to roughly 1.7% of the average staff salary and around 1.9% of the median staff salary. Taking an average of all scores in Table A4, excluding graduation rates and 8th-grade scores, I find that a 2% increase in salaries led to an approximate increase in proficiency levels of 0.05 SD (5% of a standard deviation). This is very similar to the effects observed in Biasi (2021) and slightly lower compared to the effects observed in Baron (2018).[19]

Given that not all school districts may be able to implement these salary increases through staff restructuring alone, I now estimate the additional budget required to finance a uniform $2,000 salary increase for all teaching staff. This proposed increase is slightly less than twice the average salary increment across all staff categories, yet it aligns with the upper-bound estimates for teaching staff specifically, as shown in Figures 1.3 and 1.4. I further round the amount upward to account for any associated teacher benefits. According to current data from the NJDOE New Jersey Department of Education (2024), there are approximately 600 students per school and 50 full-time teachers. At $2,000 per teacher, this would require an additional $100,000 per school, translating to an increase in per-pupil spending of roughly $170. If this amount were to be collected through increased government spending, assuming 1.4 million students attend public schools, the additional budget needed would amount to approximately $250 million. This represents roughly 2.4% of state aid for education and about 0.03% of New Jersey's Gross Domestic Product (GDP).

To put this in context, Jackson and Mackevicius (2024) finds that increasing school spending by $1,000 per pupil (sustained over four years) raises test scores by 0.03 SD and increases college enrollment by 2.8 percentage points. Based on the results, increasing per-pupil spending on teacher salaries by just $170 yields similar improvements in test scores (around 0.03 SD), suggesting a higher short-term return on investment. However, this comparison should be interpreted cautiously, as the relationship between spending and outcomes may not be strictly linear, especially at higher levels of spending. Nonetheless, under a lin-

[19]It is important to note that the effect of a decrease in teacher salaries on student achievement may not necessarily mirror that of an increase in teacher salaries.

ear assumption, increasing per-pupil spending on teacher salaries by $1,000 ($200 times 5) could potentially lead to a proficiency score increase of 0.15 SD (0.03 times 5) and raise high school graduation and college enrollment rates by around 4 (0.93 times 5) and 2 (0.44 times 5) percentage points respectively. I now explore the mechanisms through which these positive effects could have been observed.

1.10 Exploring Potential Mechanisms

Higher salaries could improve teacher performance through different mechanisms. First, increased pay may enhance productivity by raising the stakes of potential dismissal, which is in line with the efficiency wages model. Alternatively, higher salaries could improve job satisfaction, thereby motivating better performance. The second mechanism suggests that offering higher salaries enables districts to attract higher-quality teachers, either by encouraging effective teachers to switch schools or by drawing more talented individuals into the teacher labor market.

From a policy perspective, an ideal outcome would avoid districts competing for teachers, as this could exacerbate achievement gaps between schools. Moreover, a district can only afford to dismiss underperforming teachers if it has a robust pool of candidates to replace them. I now turn to an examination of what might be driving the positive results, weighing the evidence for and against each of these two mechanisms.

1.10.1 Are Teachers Switching?

The analysis here is inspired by findings from Biasi et al. (2021) and Biasi (2021). In these two papers, the authors show that granting districts control over teacher pay leads to more efficient but also more unequal teacher distribution. Efficiency improves as districts can better reward teachers for their contributions, encouraging sorting based on comparative advantage. However, inequality worsens, as teachers tend to prefer working in districts with high-achieving students. Flexible pay policies make it easier for wealthier districts to attract these teachers, thereby widening the gap. I aim to see if the data reflect a similar pattern. I provide a brief theoretical overview of this possibility in the Appendix (Section 3.6).

Ideally, the teacher dataset would include a unique identifier for each teacher to track their movement over time. However, the data lack such a variable.[20] Instead, I rely on first name, last name, and date of birth (DOB) to track teachers. This method poses challenges due to name changes (e.g., from marriage) and frequent spelling or DOB errors. To address these issues, I employ Levenshtein distance and Natural Language Processing (NLP) algorithms to match teacher names. I incorporate NLP because using only Levenshtein distance would fail to match common nicknames (e.g., Robert to Bob).[21]

Using the user-constructed teacher panel dataset offers several advantages over the previous cross-sectional dataset that lacked unique identifiers. First, the panel structure allows me to track whether teachers in treated schools had a higher probability of earning a higher degree before or after the treatment. Additionally, it enables me to assess whether these teachers switched districts or remained in their original positions. Incorporating teacher fixed effects in the event study model provides a more accurate estimation of treatment effects by accounting for unobserved teacher-specific characteristics. As a result, the estimates reflect average treatment effects at the individual teacher level, rather than district-level averages, providing more granular insights into how the salary policy influenced teacher career trajectories and educational attainment.

Patterns in Elementary Schools

I start by analyzing teacher movement patterns for Math and ELA teachers in elementary schools and find fewer than 15 occurrences of switching throughout the time period. Thus, for elementary school teachers, I do not run an event study on the number of switchers before and after the campaign. Figure 1.11 shows the results for elementary teachers.

[20] I have submitted a request to the NJDOE for data containing unique teacher identifiers.

[21] The NLP approach matched fewer than 15 individuals.

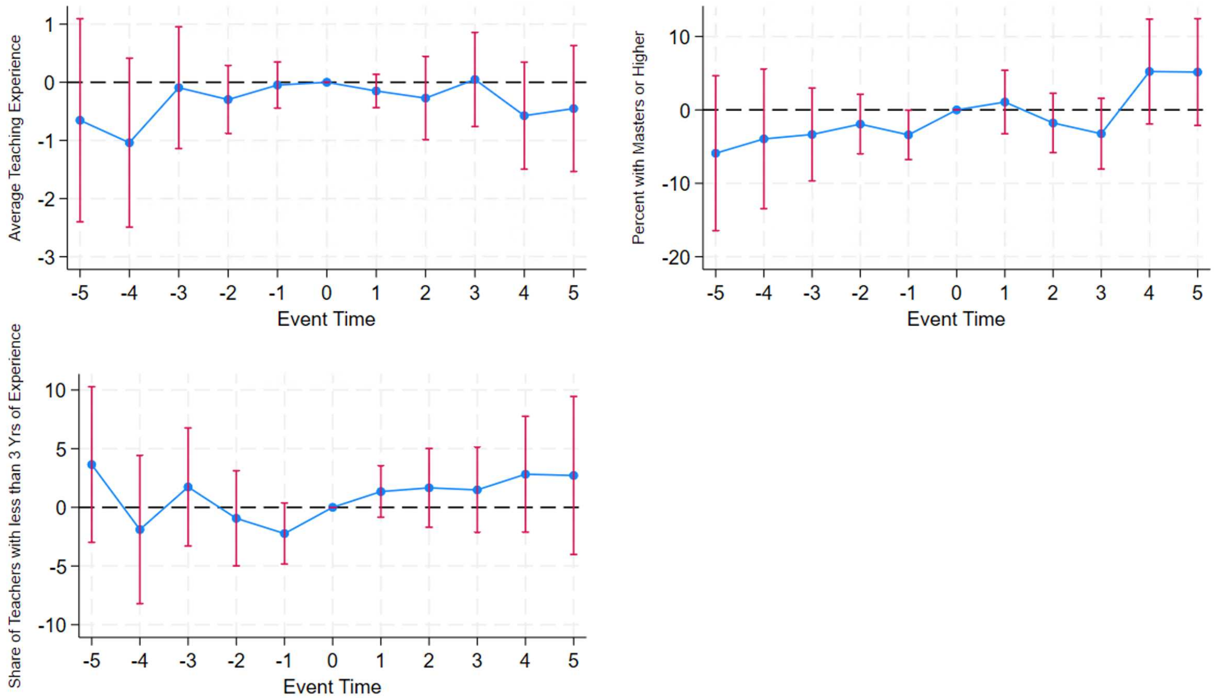


Figure 1.11: Mechanism Elementary

Note: Point estimates and 95% confidence intervals of parameters $\beta_t D_{it}$ in Equation (1.2) while also including teacher fixed effects and constructed using valid Difference-in-Differences (DiD) comparisons between treated districts and not-yet-treated districts. Average experience, total teachers with a Master's Degree or higher, and total new teachers ($<$ 3 years of experience) calculated from aggregating teacher-level data up to the school level. Treatment is continuous. The teacher dataset used incorporates NLP and Levenshtein distance.

From Figure 1.11, I can rule out any significant teacher compositional effects. Under the assumption that the teacher matching algorithm was unsuccessful, if teachers did switch post-policy, I would expect these schools to have a higher average teaching experience. However, I fail to observe this from the first figure. The top right figure then shows that the schools also did not hire more teachers with a Master's degree or higher. The bottom figure runs the event study using the share of new teachers as the outcome. Here too, I find no consistent difference in hiring patterns before or after the campaign. In summary, from these sets of event studies, I rule out that treated schools recruited “higher quality“ teachers from other schools.

Patterns in High Schools

I now analyze teacher movement patterns for Math and ELA teachers in high schools. I observe 1,620 occurrences of switching between school districts, and thus include the share of switchers in the analysis. The event study estimates are shown in Figure 1.12.

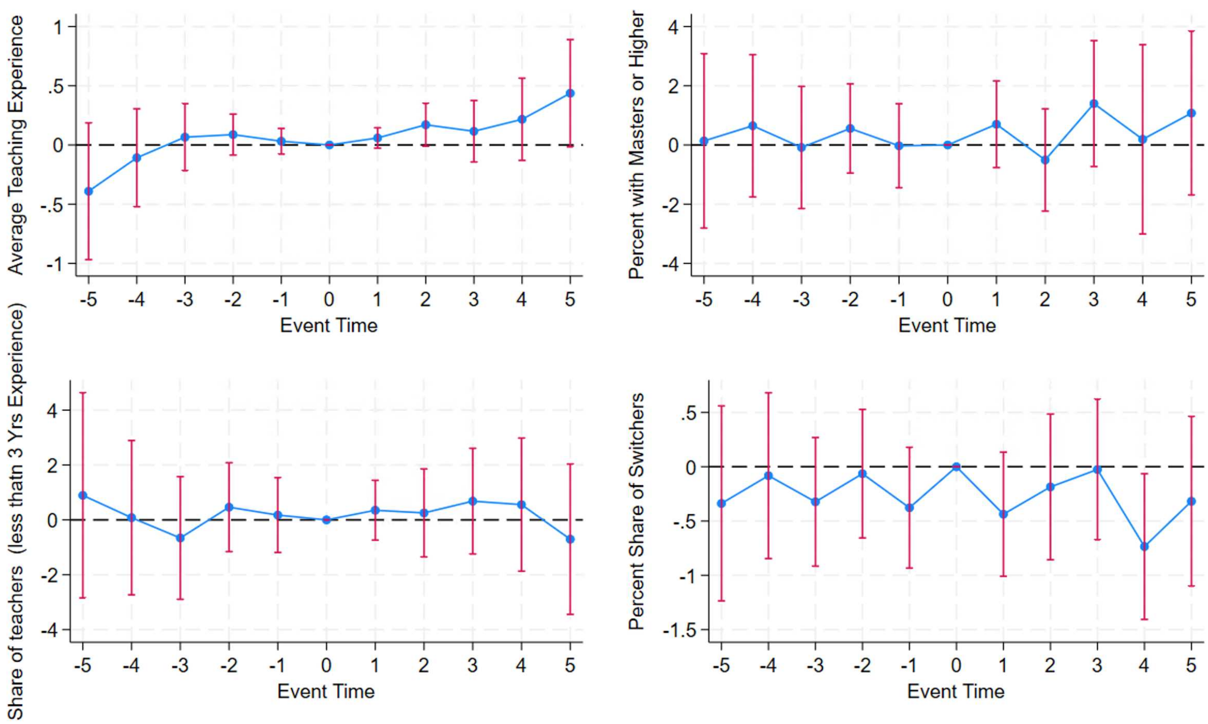


Figure 1.12: Mechanism High School

Note: Point estimates and 95% confidence intervals of parameters β_k in Equation (1.2) while also including teacher fixed effects and constructed using valid Difference-in-Differences (DiD) comparisons between treated districts and not-yet-treated districts. Average experience, total teachers with a Master’s Degree or higher, and total new teachers ($<$ 3 years of experience) calculated from aggregating teacher-level data up to the school level. Treatment is continuous. The teacher dataset used incorporates NLP and Levenshtein distance.

From Figure 1.12, I rule out significant compositional changes in the ELA and Math teachers at high schools. The average level of teaching experience remains largely unchanged, although there is a slight, statistically insignificant upward trend. The total number of teachers holding a Master's degree or higher remains stable, as does the proportion of newly hired teachers in a school. Furthermore, I find no evidence of an increased share of teachers that "switched" in treated schools. Given that the share of switchers does not increase but the average teaching experience shows an upward, albeit statistically insignificant, trend, it could be that teachers in treated schools are staying longer, resulting in lower teacher turnover rates.

1.10.2 Are teachers in treated schools performing better?

One possible mechanism through which higher salaries might lead to improved student performance is increased teacher motivation. Higher pay can attract and retain high-quality educators, foster a greater sense of professional commitment, and incentivize sustained effort in the classroom. If salary adjustments enhance teacher motivation, I should expect to see positive shifts in student outcomes in treated schools.

While I cannot directly observe teacher motivation in the dataset, I can examine whether performance improvements in treated schools coincide with abnormal declines in control schools. If salary increases are indeed driving improvements, I should observe a gradual increase in scores for treated schools, rather than a relative performance gap emerging due to a sudden and substantial drop in control schools. In such a case, the observed differences would not necessarily be driven by teacher motivation but rather by external shocks affecting the untreated schools.

Figure A9 presents the evolution of proficiency scores for four groups of schools: those treated in 2010, 2012, 2015, and 2019. Since the dataset extends only to 2019, the last group effectively represents schools that were never treated. The primary empirical strategy compares changes within each school at each treatment period, but such granular comparisons are difficult to visualize in a single graph. Instead, I categorize schools into one early-adopter group, one late-adopter group, and two mid-adopter groups to facilitate comparison.

From the graphs, I observe meaningful improvements in student proficiency levels in treated schools. Importantly, I do not observe any sudden, unexplained declines in the control schools when treated schools receive salary increases, suggesting that the observed gaps are not driven by external shocks or unrelated disruptions. This validates that the results are not threatened by systematic downward trends in untreated schools, reinforcing the interpretation that salary increases may be contributing to improved student outcomes by motivating teachers.

While these results provide strong evidence in favor of teacher motivation as a key factor, future research will aim to strengthen this claim further. Specifically, a promising next step is to obtain student-level data linked to individual teachers, allowing for a value-added analysis to directly measure teacher effectiveness. This approach would enable us to quantify how much of the observed performance improvements stem from individual teacher contributions following salary increases, offering a more precise test of the teacher motivation hypothesis.

1.10.3 Taking Stock

The results in Figures 1.11 and 1.12 allow us to rule out a key hypothesis: that improvements in student outcomes were driven by higher-quality teachers relocating to better-paying districts. This suggests that the teacher labor market dynamics in New Jersey differed from those observed in Wisconsin, where salary increases led to substantial teacher migration. The absence of similar mobility patterns in New Jersey highlights an important contextual distinction in how salary adjustments interact with local labor market structures.

A likely explanation for the lack of relocation effects lies in the nature of the salary increases themselves. Unlike performance-based pay structures that create sustained differentials across districts, these salary adjustments in New Jersey were implemented gradually and broadly, covering multiple districts over time. Because most school districts ultimately adopted similar pay schedules, the financial incentive for teachers to move was likely diminished. In contrast, the Wisconsin setting was characterized by persistent and sizable pay gaps across districts, which sustained the incentives for migration. The gradual phase-in of pay increases in New Jersey may have equalized pay scales enough to disincentivize inter-district mobility, thereby limiting the extent to which teacher sorting played a role in shaping student outcomes.

While the absence of relocation effects is clear, the role of teacher motivation remains more difficult to isolate. If salary increases improved morale and effort among existing teachers, this could plausibly explain the observed gains in student performance. Higher salaries may have increased job satisfaction, reduced attrition, and encouraged greater classroom investment, all of which could contribute to better student outcomes. However, without direct teacher-linked student performance data, it remains difficult to empirically establish motivation as the primary mechanism. Although the results are consistent with this explanation, they do not provide conclusive evidence that increased effort among existing teachers was the dominant driver of improvement.

Figure A9 provides additional insights but does not fully disentangle the competing explanations. The observed gains in treated schools could arise from multiple channels, including increased effort from incumbent teachers, changes in the composition of newly hired teachers, or broader external factors. A crucial concern in assessing these mechanisms is whether the relative performance gains in treated schools were in part driven by a deterioration in control schools rather than true improvements in treated schools. If control schools experienced an abrupt and substantial decline in scores precisely when treated schools received salary increases, the performance gap might reflect external shocks rather than any causal effect of the policy. However, the trends in Figure A9 suggest that no such systemic decline occurred. Control schools do not exhibit any sudden or anomalous drop in performance that would indicate external disruptions driving the observed gaps. The absence of such discontinuities reinforces the validity of the interpretation that salary increases played a substantive role in improving outcomes rather than merely amplifying artificial gaps caused by external shocks.

Another possible mechanism is the effect of salary increases on teacher recruitment. If both treated and non-treated districts were hiring at similar rates before and after the policy change, one potential explanation for the performance gains in treated schools is that higher salaries allowed them to attract stronger candidates. Rather than the observed improvements being solely driven by increased effort from existing teachers, the policy may have also affected the composition of the teaching workforce by drawing in higher-caliber new hires. In this scenario, salary increases may have made the profession more attractive to a broader talent pool, with higher-ability candidates disproportionately selecting into treated districts. This mechanism would be consistent with both the observed gains in student outcomes and the absence of substantial inter-district migration among incumbent teachers.

Taken together, the findings provide strong evidence against teacher relocation as a significant driver of improved student outcomes but leave open the question of whether these gains were primarily the result of increased motivation among existing teachers or changes in the composition of newly hired teachers. Distinguishing between these two channels is critical for understanding the full effects of salary increases on educational performance. While the results are suggestive, conclusive evidence would require access to teacher-linked student performance data that allows for a direct estimation of teacher value-added effects.

1.11 Conclusion

This paper examines the impact of raising teacher salaries on student outcomes by exploiting variation in the timing of district-level collective bargaining agreements that set minimum teacher salaries between \$50,000 and \$54,000. Using a staggered Difference-in-Differences (DiD) approach with a continuous treatment framework, I show that salary increases led to significant improvements in 4th-grade and high school Math and ELA scores. Additionally, I observe modest gains in high school graduation rates and college enrollment. These findings underscore the critical role of teacher compensation in shaping educational success.

The analysis reveals that these improvements are not driven by the reallocation of resources away from other expenditures or non-teaching staff. Districts achieved salary increases through modest adjustments in staffing levels while keeping per-pupil spending on non-salary components constant. This minimizes concerns about confounding effects due to broader budgetary shifts, lending confidence to the interpretation that the observed improvements stem directly from changes in teacher compensation.

This study contributes to the literature on teacher compensation in three key ways. First, I provide robust evidence that increasing teacher salaries across the board—rather than using performance-based pay schemes—can yield significant improvements in student outcomes. The observed gains of 0.05 to 0.07 standard deviations in proficiency scores are comparable to, or exceed, those documented in prior studies. Second, by ruling out teacher migration as the primary mechanism, I highlight the potential roles of teacher motivation and the recruitment of higher-quality new hires in driving these gains. However, without data directly linking teachers to student outcomes, I cannot precisely disentangle the relative contributions of these mechanisms, leaving this as a valuable avenue for future research. Finally, I offer a cost-benefit analysis that situates the findings within the broader literature on educational spending. The results suggest that modest salary increases can achieve comparable improvements in student outcomes relative to more substantial increases in overall per-pupil spending, highlighting the potential efficiency of targeted salary interventions.

While the findings are robust, they rely on several key assumptions that, if violated, could undermine the validity of the results. First, the staggered DiD framework assumes that districts adopting the $50K minimum salary floor are not systematically different from those that adopt it later. Any unobserved time-varying factors correlated with both the timing of treatment and student outcomes could bias the estimates. Although I perform robustness checks and present evidence supporting parallel pretreatment trends and comparability between early and late adopters (Table 1.4), these tests cannot entirely rule out such confounding factors. Second, the analysis assumes that treatment intensity (magnitude of salary increases) is exogenous to unobservable district characteristics. If districts with more resources or stronger teacher unions disproportionately adopted higher salary thresholds, the estimates could reflect pre-existing differences rather than causal effects. To address this concern, I confirm that pre-treatment trends in outcomes are parallel across districts with varying treatment intensities, and the results remain robust under a binary treatment assumption. While this mitigates concerns about systematic selection into treatment, I cannot fully rule out the possibility of unobservable factors driving treatment timing.

Finally, the interpretation hinges on the assumption that unmeasured shocks (statewide policy changes or economic fluctuations etc.) do not differentially affect treated and untreated districts. While I include time fixed effects to control for common shocks, district-specific heterogeneity in their effects may introduce bias. Future research could explore incorporating district-specific trends or additional robustness checks to address this limitation. The generalizability of the findings requires careful consideration. New Jersey public schools, on average, offer significantly higher teacher salaries than the national median, and its education system benefits from relatively robust funding. In states with lower baseline salaries or more constrained budgets, the marginal impact of salary increases could be larger, as teachers face greater financial stress and schools experience more acute recruitment challenges. However, structural differences in labor markets and funding mechanisms across states may limit the external validity of the findings. For example, states with weaker union presence or decentralized funding systems may find it more challenging to implement across-the-board salary increases. Future research should replicate this analysis in diverse settings to better understand how contextual factors influence the effectiveness of salary policies.

The findings have significant implications for education policy. Policymakers in states with low teacher compensation should consider salary increases as a cost-effective strategy for improving student outcomes. By demonstrating that even modest, across-the-board salary increases yield measurable improvements, this study underscores the value of prioritizing teacher pay in resource allocation decisions. However, careful attention should be paid to potential trade-offs, including impacts on staffing levels and non-teaching resources.

In conclusion, this paper provides robust evidence that increasing teacher salaries has a substantial and positive impact on student academic outcomes. Within the broader context of school finance, the findings suggest that targeted investments in teacher compensation may generate greater returns than more diffused spending strategies. By improving student achievement through targeted salary increases, education systems can enhance both academic outcomes and long-term life prospects for students, reinforcing the critical role of teachers in shaping societal progress.

Chapter 2

Bridging the Gap to Access? Impact of Pre-Collegiate Summer Program on College Outcomes

U.S. colleges face persistent inequities—not only in who enrolls, but in who ultimately graduates. While institutions have expanded efforts to diversify incoming classes, students from historically underserved backgrounds, including those who are first-generation and low-income, continue to face several barriers which affect their rate of degree completion. This paper evaluates the causal impact of the Bridge Scholars Program (BSP), an eight-week pre-collegiate summer initiative at a large public land-grant college[22] designed to support such students in their transition to college. Leveraging a fuzzy regression discontinuity design (FRDD) within a local randomization inference framework, I find that BSP substantially improves first-year completion and second-year persistence. While participants receive a $2,500 annual scholarship, comparative analyses of similar financial aid programs suggest that the bundled nature of BSP—including structured summer programming, academic mentoring, and peer support—is central to its effectiveness. These findings underscore the importance of holistic interventions that address both financial and non-financial barriers to student success.

[22] This research was reviewed and approved by the university's Institutional Review Board (IRB), protocol number 4392. I would like to thank Nicole Ross for all her help throughout this process.

2.1 Introduction and Motivation for the Study

Higher education in the United States continues to face persistent inequities in access and completion, particularly among students from low-income, first-generation, and underrepresented racial/ethnic backgrounds. These disparities reflect a combination of structural inequalities and individual-level barriers, including differences in K–12 preparation, affordability, social capital, and institutional support. Over the past several decades, while overall college enrollment has expanded, the gap in college completion between students from high- and low-income families has widened considerably (Bailey & Dynarski, 2011; Belley & Lochner, 2007). Only 13% of students from the lowest income quartile complete a bachelor's degree by age 24, compared to 64% from the highest income quartile (Cahalan et al., 2021).

A substantial literature has examined the effectiveness of financial aid in promoting college access and persistence, finding that while generous grants can improve enrollment, aid alone is often insufficient to ensure degree completion (Dynarski, 2003; L. C. Page & Scott-Clayton, 2016). Recent work has highlighted the role of non-financial barriers—such as inadequate information, poor academic preparation, limited advising, and psychological frictions—as critical impediments to student success, particularly for disadvantaged students (Dynarski et al., 2023). Interventions that combine financial aid with personalized guidance and academic support appear to be most effective in promoting long-term educational attainment (Castleman & Page, 2015; Oreopoulos et al., 2017).

In this context, summer bridge programs have emerged as a common institutional strategy aimed at easing the transition to college. These programs typically target academically underprepared or underserved students and offer structured academic coursework, skill-building workshops, mentoring, and early exposure to college life. The premise is that a well-supported entry into college can foster social integration, promote academic momentum and early access to campus resources such as the library, academic advising etc. Nationally, by the early 2010s, approximately 13% of four-year institutions had implemented such initiatives (Barnett et al., 2012). Evaluating these programs is essential for two reasons: first, to determine whether they improve academic and retention outcomes; and second, to guide institutional decisions about scaling and resource allocation, particularly when programs serve small cohorts or demand substantial investment.

Despite their popularity, empirical evidence on the effectiveness of summer bridge programs remains mixed (Barnett et al., 2012; Murphy et al., 2010). While some studies report short-term academic gains, such as higher GPAs or greater credit accumulation, longer-term impacts on persistence and graduation are less consistent. While there has been a somewhat growing interest in evaluating summer bridge programs, few evaluations of bridge programs meet the methodological standards set by the What Works Clearinghouse (WWC), an initiative of the Institute of Education Sciences (IES) that evaluates education interventions based on transparent criteria for causal inference and replicability. According to WWC, the existing evidence base suffers from small samples, design weaknesses, and limited external validity (What Works Clearinghouse (WWC), 2024). In their 2016 report analyzing summer bridge programs up to that date, WWC's report showed that only the two aforementioned studies met the WWC group standard.[23]

This paper contributes to the literature by evaluating a comprehensive intervention: the Bridge Scholars Program (BSP). BSP is an eight-week residential summer bridge program that combines intensive academic support with a yearly $2,500 scholarship each year they stay enrolled at the college in question (for upto 4 years of college). It targets incoming low-income, first-generation students and includes credit-bearing coursework, advising, mentoring, and structured community engagement.

To estimate the causal impact of BSP participation, I implement a fuzzy regression discontinuity design (FRDD) exploiting a program eligibility threshold based on Expected Family Contribution (EFC). This design allows for quasi-random assignment around the

[23]Other papers close to summer bridge evaluation includes Cabrera et al. (Cabrera et al., 2013) which found positive associations between participation in the New Start Summer Program and student success, and Kitchen et al. (Kitchen et al., 2018) which reported positive effects of STEM-focused bridge programming. Both studies, however, use designs that may not fully account for selection dynamics.

EFC cutoff, offering credible estimates of the program's impact on academic performance and retention for students near the eligibility margin. The findings indicate significant positive effects of BSP participation on college persistence which I define as the decision by the student to remain enrolled at the college. Beyond the direct program effects, this paper also explores the mechanism through which the positive effects could have been observed. Findings suggest that the mixture of summer support and financial assistance is the primary driver of improved persistence.

2.2 Overview of the Bridge Scholars and Data Description

The Bridge Scholars Program (hereby BSP or Bridge) is an eight-week summer initiative designed to support first-generation and low-income students during their transition into college. Participants enroll in credit-bearing courses, reside in a structured residential community, and engage in academic workshops and social activities intended to foster academic readiness, build community, and connect students to campus resources. BSP students also receive a $2,500 annual scholarship renewable for four years, contingent upon continued participation and satisfactory academic progress. The program's dual emphasis on academic preparation and financial assistance directly targets the most common barriers faced by underserved student populations.

Once students have been accepted to the college in question, eligible students receive an invitation to apply for the summer program. The outreach is based on criteria such as first-generation status, income, school district area, and other relevant indicators of social and economic disadvantage. The application materials make it clear that the program is intended for students from underrepresented backgrounds and does not require academic credentials like high school GPA. Since Fall 2016, high school GPA has not been a factor for consideration in the application process and the selection committee has instead placed stronger emphasis on first-generation status and limited-income indicators, particularly Expected Family Contribution (EFC).

The dataset I use for my analysis includes student-level data from a large public land-grant university in the US, covering all undergraduates who enrolled from Fall 2013 through Fall 2022, with academic performance tracked through January 2024. I exclude cohorts prior to 2016 because the Bridge Scholars Program (BSP) was substantively different in earlier years and restrict the analysis to students who applied to the BSP. Similarly, the Fall 2023 cohort is excluded because, at the time of data collection, only one semester of academic outcomes was available.

For the main analysis, I limit the sample to students who applied to Bridge Scholars program and subsequently enrolled at the college. BSP students are those who applied to the program and were selected, while non-BSP students are those who applied but were not selected to participate in the summer program. Both groups ultimately attend the college. This structure mitigates some potential selection bias by focusing exclusively on applicants, ensuring that both treated and control groups expressed interest in participating in the program. Table 2.1 presents the distribution of BSP and non-BSP students across the six included cohorts.

Table 2.1: Bridge Scholars Program Participation by Entering Cohort

Cohort (Fall)	Non-BSP	BSP	Total
FA17	5	36	41
FA18	9	36	45
FA19	24	38	62
FA20	24	37	61
FA21	25	39	64
FA22	56	81	137
Total	143	267	410

Note: As Table 1 shows, BSP participation has grown over time, with the number of students in the program increasing substantially in more recent cohorts.

As Table 2.1 shows, BSP participation has grown over time, with the number of students in the program increasing substantially in more recent cohorts. Table 2.2 summarizes the demographic and academic characteristics of BSP and non-BSP students across these cohorts. Among BSP participants, 95% are first-generation college students, 94% come from limited-income backgrounds as defined by Pell Grant eligibility and low EFC, and 81% identify as Hispanic/Latino. Crucially, the two groups do not differ meaningfully in high school GPA, underscoring that academic preparation is not a selection criterion.

Table 2.2: Applicant Characteristics by BSP Status

Characteristic	Non-BSP	BSP	Total
Total Sample	143 (35%)	267 (65%)	410 (100%)
Race/Ethnicity			
Asian	7 (5%)	7 (3%)	14 (3%)
Black	6 (4%)	10 (4%)	16 (4%)
Hispanic/Latino	99 (69%)	217 (81%)	316 (77%)
White	23 (16%)	26 (10%)	49 (12%)
Other/Multi-Racial	8 (6%)	7 (3%)	15 (4%)
Income Status			
Limited Income	15 (10%)	252 (94%)	267 (65%)
Not Limited Income	128 (90%)	15 (6%)	143 (35%)
First-Generation Status			
First-Generation	125 (87%)	255 (95%)	380 (93%)
Not First-Generation	18 (13%)	12 (5%)	30 (7%)
High School GPA			
Mean GPA	3.72	3.66	3.68
Standard Deviation	0.43	0.48	0.46

Note: Percentages within each column (Non-BSP, BSP, and Total) sum to 100% within each section. "Other" includes multi-racial students or those with unspecified ethnicity. GPA reflects self-reported high school GPA. International students are not eligible for BSP.

In terms of Expected Family Contribution (EFC), the majority of applicants to the program report an EFC below $20,000 (Figure 2.1). On average, applicants had an EFC of approximately $7,000 (SD = $11,285). In contrast, the average EFC among the broader college population (excluding applicants) is approximately \$32,000 (SD= \$55,997). These figures suggest that the applicants were from significantly lower-income backgrounds compared to the general student body. Given this demographic difference, it is plausible that Bridge program (BSP) participants receive more financial aid than non-participants.

Table 2.3 presents summary statistics on financial aid and scholarships received by BSP and non-BSP students, excluding the BSP-specific \$2,500 annual scholarship awarded upon successful program completion. BSP participants receive significantly more financial support overall, with higher average amounts of both need-based aid and merit-based scholarships in their first year. On average, BSP students receive \$15,491 in first-year need-based aid, compared to \$11,482 for non-participants. Similarly, they receive \$6,095 in other scholarships, exceeding the \$5,278 average received by their non-BSP peers. This pattern persists into the second year, with BSP students continuing to receive more need-based aid.

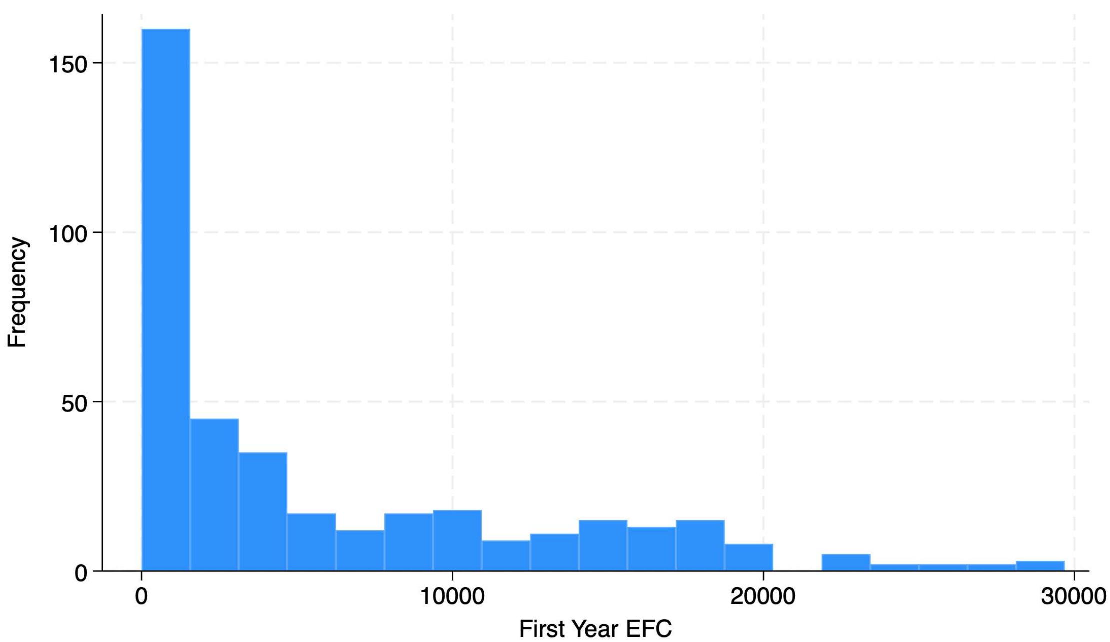


Figure 2.1: Distribution of First-Year Expected Family Contribution (EFC)

Note: Distribution shown for Bridge program applicants; EFC values above \$30,000 are omitted.

Table 2.3: Summary Statistics for Financial Aid and Scholarships (Excluding BSP Scholarship)

Group	1st Yr Need-Based	2nd Yr Need-Based	1st Yr Scholarship
Non-BSP (Mean)	$11,481.64	$5,981.30	$5,278.17
BSP (Mean)	$15,491.44	$7,663.19	$6,095.46

Note: BSP-specific scholarship ($2,500) is excluded. Higher average aid for BSP students reflects targeting of financially needy populations.

The observed financial disparity raises a valid concern regarding potential endogeneity: students receiving greater aid may systematically differ from others in ways that also predict better academic outcomes. In particular, if BSP participants benefit from more generous financial aid packages beyond the BSP scholarship itself, isolating the program's unique impact becomes empirically challenging. Importantly, as shown later in the analysis, there is little evidence that these additional scholarships account for the observed differences in performance and persistence. Within the sample used to estimate program effects, I find no meaningful differences in covariates—including financial aid—between students just above and just below the program eligibility threshold. This reinforces the credibility of the identification strategy used to assess the impact of BSP.

2.3 Motivating Question and Preliminary Results: Ordinary Least Squares (OLS) and Nearest Neighbor Matching

In this paper, I focus primarily on persistence—whether students remain enrolled into subsequent semesters—as the main outcome of interest. This measure is important because first-generation and limited-income students are disproportionately at risk of early college departure. Persistence reflects not only academic engagement but also the broader social and structural challenges these students face in higher education environments (Roksa & Kinsley, 2019; Stewart et al., 2015).

In addition to persistence, I examine the total number of successfully attempted credits during the first academic year. Credit accumulation is a critical indicator of academic momentum and degree progress. I do not place primary emphasis on GPA because academic grading norms can vary substantially across departments and majors and I do not have enough observations to estimate program impacts by college major choices. Furthermore, since many students in the sample did attend atleast some semesters during the pandemic, comparing GPA's across years might result in biased estimates given the possibility of grade inflation during the pandemic semesters. Lastly, I explore STEM enrollment (choosing majors enlisted as pertaining to areas of study focused in Science, Technology, Engineering and Mathematics) outcomes using a probit model. However, this analysis is exploratory; the Bridge Scholars Program was not explicitly designed to increase STEM participation, and therefore these results are presented as secondary. Given this focus on persistence and credit attainment, the covariates included in my models are selected based on prior research identifying key predictors of college retention among disadvantaged student populations (Roksa & Kinsley, 2019; Stewart et al., 2015).

Before introducing the main identification strategy, I begin with a set of baseline Ordinary Least Squares (OLS) regressions. Table 2.4 presents coefficients from models regressing academic outcomes—including first and second semester GPA, second-year GPA, and total first-year credits—on BSP participation and relevant controls.[24] As expected, high school GPA is consistently predictive of subsequent college GPA. BSP participation is associated with a statistically significant improvement in first-term GPA, but this effect diminishes in the second semester, and no significant differences are observed by the second year. BSP

[24] I define a student's gender as binary (Male or Female) because this is the level at which the institutional data is collected.

participants also appear to earn slightly fewer total credits in their first year than applicants who were not selected into the program.

Table 2.4: OLS Estimates for Academic Outcomes

Covariates	**FA1 GPA**	**SP1 GPA**	**FA2 GPA**	**1st Yr Credits**
HS GPA	0.606***	0.578***	0.455***	2.826***
	(0.089)	(0.086)	(0.073)	(0.766)
BSP Cohort	0.224***	0.124*	0.077	-1.148*
	(0.082)	(0.073)	(0.070)	(0.656)
Merit Aid ($1000s)	0.020**	0.013*	0.012*	0.060
	(0.008)	(0.007)	(0.007)	(0.064)
Gender: Male	-0.034	-0.057	-0.087	-0.227
	(0.081)	(0.072)	(0.067)	(0.642)
Constant	0.469	0.719**	1.154***	17.884***
	(0.350)	(0.335)	(0.293)	(2.994)
N	404	380	342	380
R^2	0.186	0.177	0.160	0.094

Note: Each column presents results from a separate OLS regression. All models include fixed effects for cohort entry year (wave), which are omitted from display. Merit aid is scaled in thousands of dollars. Robust standard errors are reported in parentheses.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table 2.5 shows probit regression estimates for three binary outcomes: persistence to the second fall semester, completion of the second spring semester, and enrollment in a STEM major by the second fall. The model includes controls for high school GPA, college GPA at the prior semester, gender, cohort fixed effects, and merit aid (measured in thousands). Bridge participation is not significantly associated with persistence, and the estimates for academic predictors such as GPA are strong and positive. Male students are significantly more likely to be enrolled in a STEM major in the second year, and higher merit aid is also associated with higher likelihood of STEM participation. However, the effect of BSP on STEM enrollment is small and not statistically significant.

Table 2.5: Probit Estimates for Persistence and STEM Outcomes

Covariates	**Persisted to FA2**	**Completed SP2**	**STEM in FA2**
Cumulative GPA (previous semester)	0.509***	2.292***	-0.147
	(0.132)	(0.528)	(0.125)
High School GPA	0.253	0.285	0.351*
	(0.261)	(0.353)	(0.195)
BSP Cohort	0.199	0.389	0.065
	(0.204)	(0.443)	(0.158)
Merit Aid ($1,000s)	-0.004	0.074	0.041***
	(0.023)	(0.057)	(0.015)
Gender: Male	-0.070	-0.028	0.463***
	(0.207)	(0.444)	(0.151)
Constant	-0.867	-5.213***	-1.896**
	(0.993)	(1.633)	(0.740)
Cohort Year FE	Yes	Yes	Yes
Observations	380	307	339
Pseudo R^2	0.117	0.577	0.053

Note: Dependent variables are binary indicators for persistence to the second fall (FA2), completion of the second spring (SP2), and STEM major enrollment in FA2. Merit aid is scaled in thousands of dollars. Robust standard errors are in parentheses.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

To bolster robustness and account for possible differences in covariate distributions, I complement the OLS analysis with a nearest-neighbor matching (NNM) approach. Following Abadie et al. (2004, 2011), I estimate the Average Treatment Effect on the Treated (ATET) for each outcome and for this process, the sample size includes the entire college population and not just the applicant group.

Specifically, each BSP participant is matched with the most similar non-applicant based on Mahalanobis distance across a set of covariates: high school GPA, race/ethnicity, gender, first-generation status, EFC, and entry year (cohort). To reiterate, this sample includes all full-time college students during that period.

Let $T_i \in \{0,1\}$ denote treatment (BSP participation), and let Y_i^T and Y_i^C represent outcomes for treated and control students, respectively. The matching estimator is:

$$\widehat{\text{ATET}} = \frac{1}{N_T} \sum_{i:T_i=1} (Y_i^T - Y_{j(i)}^C)$$

where $j(i)$ is the matched non-treated control for treated unit i. I also apply the bias-corrected adjustment proposed by Abadie et al. (2011) to account for imperfect matches on continuous covariates.

As shown in Table 2.6, BSP participants are 8–10 percentage points more likely to persist through the second year than matched controls. GPA effects are mixed: no advantage in the first term, but a 0.10 point disadvantage in the second semester. STEM-related outcomes remain unchanged.

Table 2.6: ATET Estimates: Bridge Students vs. Similar Non-Bridge Students (non Applicants)

Outcome Variable	Coefficient	p-value	[95% Conf. Interval]	Min-Max Matches
SP1 GPA	0.0296	0.494	[-0.0551, 0.1143]	15 - 53
SP2 GPA	-0.0999	0.011	[-0.1766, -0.0233]	15 - 47
STEM 1st Yr.	-0.0235	0.446	[-0.0840, 0.0369]	15 - 53
STEM 2nd Yr.	0.0216	0.517	[-0.0437, 0.0869]	15 - 47
Persisted to FA2	0.0818	0.000	[0.0406, 0.1231]	15 - 57
Persisted to SP2	0.0954	0.000	[0.0495, 0.1413]	15 - 57

Note: Matching results comparing Bridge participants with similar non-Bridge students. This comparison does not include students that applied to the Bridge program but were not accepted. Persistence outcomes show significant positive effects, while GPA improvements diminish over time.

Given that selection into the Bridge Scholars Program (BSP) is not random, a major concern is the possibility of selection bias: students who were ultimately accepted to participate in BSP may be more motivated, academically prepared, or otherwise distinct from their non-selected peers. Table 2.7 provides suggestive evidence that such bias may exist. Specifically, students who applied to BSP but were not admitted perform significantly worse in terms of GPA than those who were selected. When both BSP participants and non-selected applicants are compared to the broader college student population, it appears that non-selected applicants may be systematically lower performing. However, it is unclear whether these differences arise because these students lacked access to BSP support or because they were intrinsically lower achieving and therefore less likely to be selected in the first place. This ambiguity is evident across both OLS and nearest-neighbor matching (NNM) estimates which fail to disentangle selection effects from treatment effects. These patterns underscore the risks of relying on naive comparisons: observed differences in academic outcomes may reflect pre-existing disparities rather than the causal impact of BSP. Thus, I employ a quasi-experimental approach—namely a local randomization regression discontinuity design–to address concerns about endogeneity and selection bias and to get credible estimates of the program effects.

Table 2.7: ATET Estimates: Applicants Not Selected vs. Non-Applicants

Outcome Variable	Coefficient	p-value	[95% Conf. Interval]	Min-Max Matches
SP1 GPA	-0.1686	0.011	[-0.2984, -0.0387]	15 - 36
SP2 GPA	-0.2279	0.000	[-0.3360, -0.1199]	15 - 29
STEM 1st Yr.	-0.0149	0.751	[-0.1073, 0.0774]	15 - 36
STEM 2nd Yr.	-0.0267	0.587	[-0.1231, 0.0696]	15 - 29
Persisted to FA2	0.0100	0.782	[-0.0606, 0.0806]	15 - 40
Persisted to SP2	-0.0178	0.651	[-0.0949, 0.0593]	15 - 40

Note: Matching results for students who applied but were not selected, compared to non-applicants. No significant effects on persistence or STEM enrollment.

2.4 Empirical Approach: Fuzzy RDD under Local Randomization

Although BSP admission is holistic, program participation is primarily governed by a student's Expected Family Contribution (EFC). As a reminder to the readers, since 2016, high school GPA has not factored directly into selection decisions, which allows EFC to serve as a credible running variable in a fuzzy regression discontinuity design (RDD), with $16,000 functioning as the de facto eligibility threshold.

Figure 2.2 illustrates a clear discontinuity in the probability of program participation at the $16,000 cutoff. The relationship is not perfectly sharp—some students with EFCs above the threshold were still admitted, while some below it were not. These deviations may reflect non-financial considerations such as weak personal statements, unexpected changes in family financial status, or administrative discretion. Nonetheless, the observed jump in treatment probability at the cutoff confirms the presence of a strong first-stage relationship necessary for implementing a fuzzy RDD framework.

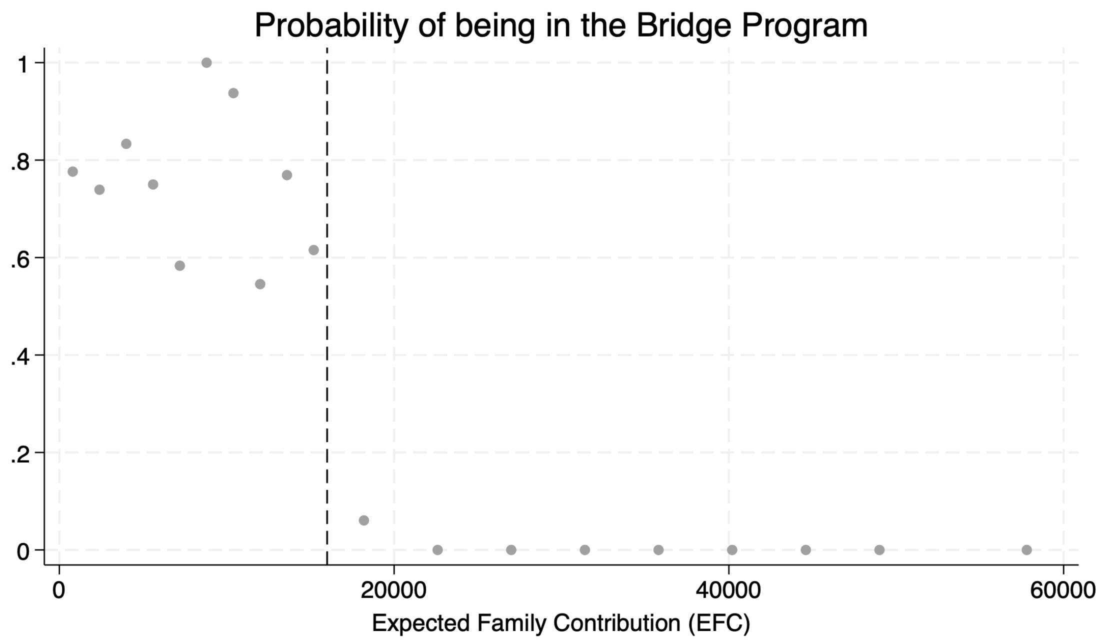


Figure 2.2: First-Stage Discontinuity in BSP Assignment

Note: Graph illustrates the discontinuous probability of BSP assignment at the $16,000 EFC threshold, confirming a strong first-stage relationship necessary for fuzzy regression discontinuity analysis.

Despite this strong first-stage relationship, implementing a conventional continuity-based RDD is challenging. As shown in Table 2.8, the sample includes only 410 observations, with substantial imbalance on either side of the cutoff. The effective sample size near the threshold is small, reducing statistical power for detecting treatment effects, even when they are sizable.

Table 2.8: Power Test Summary: Continuous RDD Framework at \$16,000 EFC Cutoff

Parameter	**Value**	**Description**
Cutoff	\$16,000	EFC threshold for BSP prioritization
Observations (Left of Cutoff)	338	Raw sample size left of cutoff
Observations (Right of Cutoff)	63	Raw sample size right of cutoff
Effective Obs (Left)	67	Weighted obs within optimal bandwidth
Effective Obs (Right)	36	Weighted obs within optimal bandwidth
Order of Local Polynomial	1	Linear regression at cutoff
Bandwidth (h)	7,862.89	Optimal bandwidth selection
Kernel	Triangular	Weighting scheme
Power Against Effect Size (τ)		**Power (%)**
$0.2 \times \tau$		11.0
$0.5 \times \tau$		27.6
$0.8 \times \tau$		44.2
$1.0 \times \tau$		55.2

Note: Power estimates are based on simulations under the continuous RDD framework. Despite a strong first-stage discontinuity, the limited number of effective observations near the threshold curtails statistical power, particularly for detecting moderate-sized effects. This limitation motivates the use of a local randomization regression discontinuity approach in the next section.

In response, I adopt a local randomization fuzzy RDD to causally evaluate the effects of BSP. This framework overcomes many of the limitations inherent in continuity-based designs by allowing for valid inference in small samples, formalized by Branson and Mealli (2018) and Cattaneo et al. (2024).

Despite its growing formal foundations, the local randomization framework remains underutilized in applied settings, particularly within economics and education. As Branson and Mealli (2018) note, few empirical studies have implemented this approach, even though it offers a valid and transparent identification strategy when continuity assumptions or large samples are lacking.[25] My contribution here is thus both substantive and methodological: by applying local randomization to a fuzzy RDD context, I show that causal inference is feasible and interpretable even in the face of real-world design constraints.

While selection is not deterministic, students just below the cutoff are substantially more likely to be admitted. Let X_i denote EFC, Z_i indicate assignment below the threshold, D_i the indicator for BSP participation, and Y_i the outcome of interest. Within a local window W, I estimate the causal effect of treatment on the treated via two-stage least squares (TSLS):

$$D_i = \pi_0 + \pi_1 Z_i + \nu_i, \tag{2.1}$$

$$Y_i = \alpha + \tau D_i + \epsilon_i, \tag{2.2}$$

where π_1 captures the first-stage jump in participation and τ is the local average treatment effect (LATE) for compliers near the threshold.

I emphasize TSLS over the simpler intention-to-treat (ITT) framework because it isolates the causal effect of actual program participation, not just eligibility. While the ITT—which compares mean outcomes between $Z_i = 1$ and $Z_i = 0$—is estimated and presented in the appendix, it conflates selection with treatment. Moreover, in settings with substantial non-compliance, TSLS yields the more policy-relevant estimand.

Under the local randomization framework, I assume that within a narrow window around the cutoff, treatment assignment is as good as random, conditional on observed covariates. The estimand is no longer a limit (as in continuity-based RDD), but a finite-sample average treatment effect over the units in window W:

[25] For applied studies using local randomization, see Li et al. (2015) on university grants in Italy and Cunial (2021) on solar subsidies in rural Colombia.

$$\tau_W = \frac{1}{|W|} \sum_{i \in W} \left[Y_i(1) - Y_i(0) \right], \tag{2.3}$$

with inference proceeding via either Fisherian or Neymanian logics. Fisherian inference tests a sharp null hypothesis ($Y_i(1) = Y_i(0)$ for all i) using exact permutation-based p-values, and is valid in finite samples without distributional assumptions. Neyman-style inference, by contrast, focuses on estimating average effects and constructing confidence intervals under large-sample approximations.

The `rdrandinf` STATA package used in this paper offers both modes of inference. For the ITT effect, Fisherian p-values are available by default through randomization tests. However, for TSLS, `rdrandinf` reports only large-sample p-values, because the LATE estimator is nonlinear (a ratio) and not well-suited to permutation-based tests (Cattaneo et al., 2016, 2024). Nonetheless, I obtain confidence intervals for the TSLS estimates by inverting the estimated variance of the Wald ratio under asymptotic normality. These are reported throughout the main results.

Although exact p-values are unavailable for the LATE under local randomization, the primary inferential focus of this paper is on confidence intervals rather than on precise point estimates. This emphasis is consistent with the design-based philosophy underpinning local randomization, where the goal is to detect whether plausible treatment effects exist within an interpretable, covariate-balanced window and not more so on obtaining precise program estimates. As Branson and Mealli (2018) and Cattaneo et al. (2024) emphasize, even under imperfect compliance or limited power, the width and inclusion of the confidence interval offer meaningful insight into the presence and magnitude of plausible program effects.

Without wanting to get ahead of the results, I find that Fisherian inference for the ITT and Neyman-style inference for the TSLS-based LATE yield broadly consistent conclusions. While the underlying inferential frameworks differ—Fisherian inference tests a sharp null hypothesis using permutation-based p-values, whereas Neyman inference relies on large-sample approximations to construct confidence intervals—both methods point to similar results. This convergence across inferential strategies enhances the credibility of the findings, especially given the small sample sizes and partial compliance typical of educational interventions.

Window selection follows the algorithmic procedure in Cattaneo et al. (2024), in which I iteratively expand the local window until covariate balance across treatment and control groups begins to deteriorate. At each step, I assess balance using randomization-based statistics such as the Mahalanobis distance and joint difference-in-means tests, consistent with best practices in the local randomization literature. Although I do not present manipulation tests such as the McCrary density test here, these are reported later in the paper and provide additional support for the integrity of the running variable.

My design supports valid causal inference even in the presence of modest sample sizes and fuzzy compliance. By interpreting the TSLS estimate as a finite-population LATE defined over units within the selected window, and by constructing robust confidence intervals around this estimate, I am able to quantify the impact of BSP participation on academic persistence and early credit accumulation among students near the margin of eligibility. Recent work has helped clarify when regression discontinuity designs perform well in small samples. Swartzentruber and Kaizar (2022) introduce a density-adjusted measure of effective sample size and show that valid inference is possible even with small samples, as long as enough observations lie close to the cutoff. I follow the data-driven window selection methods of Cattaneo et al. (2024) to ensure that density near the $16,000 EFC threshold is adequate, helping to avoid this concern. A different issue, raised by Litschwartz (2022), is that local randomization RDDs can be biased when the running variable is a noisy signal of an unobserved trait that strongly predicts outcomes—such as standardized test scores used as imperfect measures of academic ability. In my context, however, the running variable is Expected Family Contribution (EFC), a precise administrative measure used for determining financial aid eligibility. While EFC reflects socioeconomic status, it is not a proxy for academic preparedness and is only weakly predictive of college GPA. Moreover, I include high school GPA—a far stronger predictor of academic outcomes—as a control variable. Therefore, I argue that the type of bias that the working paper describes is unlikely to apply to this study.

While not the central contribution of this paper, it is worth emphasizing that local randomization regression discontinuity designs offer not only credible small-sample inference but also a potentially stronger foundation for generalization beyond the estimation window. As Branson and Mealli (2018) argues, because the estimand in local randomization is defined over a finite set of observed units—rather than as a limit at an infinitesimal neighborhood of the cutoff as in continuity-based RDD—the underlying assumptions required to extrapolate to adjacent subpopulations may be more empirically transparent and policy-relevant than the estimates observed from a continuous RDD approach. Specifically, when effects are shown to be stable across multiple, overlapping, covariate-balanced windows, the resulting evidence can support broader inference across a meaningful portion of the running variable's range. Although extrapolation is not the primary focus of this paper, the structure of the local RDD design adopted here is especially well suited for informing policy decisions that apply to a wider range of economically vulnerable students—not merely those infinitesimally close to the $16,000 EFC threshold.

2.5 Main Results

The outcomes examined include measures of persistence, credit accumulation, GPA, and STEM classification. Data for completed credits and STEM major designation are only recorded at the end of each academic term. As a result, students who withdrew or dropped out are missing from these outcomes, leading to a lower number of observations relative to persistence measures. The GPA variable reflects the most recent cumulative college GPA on record at the time of student departure, not semester-specific GPA. For example, if a student dropped out in March (Spring), I use the GPA in their Fall semester.

Table 2.9 presents the core Two-Stage Least Squares (TSLS) estimates from a fuzzy local randomization regression discontinuity design (RDD) centered at the BSP eligibility cutoff of \$16,000 in Expected Family Contribution (EFC). The identification window is selected using covariate balance across high school GPA, first-year need-based aid, and merit-based aid (excluding BSP awards), resulting in a symmetric local bandwidth of [5293, 26707] and an effective sample of 135 students (92 below and 43 above the cutoff). End of Year Completed Credits and End of Term STEM classification is not recorded for students who dropped out of college or withdrew during the first year which results in a lower number of observation for those outcomes.

Table 2.9: TSLS Estimates from Fuzzy Local Randomization RDD (EFC Cutoff = \$16,000)

Outcome	**TSLS Estimate**	**95% CI**	**Obs (Below / Above)**
Persistence (Complete 1st Yr)	0.469***	[0.238, 0.699]	92 / 43
Complete 2nd Year	0.301**	[0.059, 0.543]	92 / 43
First-Year Completed Credits	1.561	[-2.566, 5.689]	84 / 34
STEM Major (Spring Yr 1)	-0.062	[-0.346, 0.222]	84 / 34
Standardized 1st Yr. GPA	0.025	[-0.009, 0.059]	92 / 43

Note: Estimates are based on Two-Stage Least Squares (TSLS) under a fuzzy local randomization design centered at an EFC cutoff of \$16,000. The data driven local window is [5293, 26707] based on covariate balance using high school GPA, first-year need-based aid, and merit scholarships excluding BSP-specific awards.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

The results in Table 2.9 indicate a statistically and substantively significant effect of BSP eligibility on short-run persistence outcomes. Students just below the \$16,000 EFC threshold are estimated to be 46.9 percentage points more likely to complete their first year of college ($p < 0.01$), and 30.1 percentage points more likely to persist through the second year ($p < 0.05$), relative to their counterparts just above the cutoff. Importantly, the 95%

confidence intervals provide compelling evidence that these effects are not only statistically significant but also meaningfully large even under conservative interpretation. For example, the lower bound for first-year completion is 23.8 percentage points and 5.9 percentage points for second-year persistence—suggesting that even in the most pessimistic case consistent with the estimates, BSP eligibility yields non-trivial improvements in student retention. By contrast, the program's estimated effect on first-year credit accumulation is positive but statistically imprecise. The confidence interval ranges from -2.6 to 5.7 credits, reflecting considerable uncertainty. Finally, I observe no statistically meaningful impacts on STEM major enrollment or GPA during the first year; both estimates are small in magnitude and their confidence intervals include zero.

Table 2.10: TSLS Estimates by COVID Cohort Status (Fuzzy Local Randomization RDD, EFC Cutoff = $16,000)

Sample and Specification	TSLS Estimate	95% CI	Obs (Below / Above)	Window [EFC]
COVID Cohort Only				
Complete 1st Year	-0.083	[-0.253, 0.087]	17 / 7	[5293, 26707]
Complete 2nd Year	-0.167	[-0.398, 0.064]	17 / 7	[5293, 26707]
Non-COVID Cohorts				
Complete 1st Year	0.570^{***}	[0.315, 0.824]	75 / 36	[5293, 26707]
Complete 2nd Year	0.384^{***}	[0.119, 0.649]	75 / 36	[5293, 26707]

Note: Estimates based on fuzzy local randomization regression discontinuity designs centered at an Expected Family Contribution (EFC) cutoff of $16,000. Models use Two-Stage Least Squares (TSLS) to estimate the local average treatment effect of BSP eligibility. The first two panels use a consistent user-specified window of [5293, 26707]. The final panel employs an optimal covariate-balanced window selected from among non-COVID cohorts. Expanded covariates include high school GPA, need-based aid, merit aid (excluding BSP award), gender, STEM interest at entry, and age at college start. Confidence intervals are based on asymptotic inference.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Outcomes related to academic performance and field-of-study orientation—such as standardized first-year GPA and spring-term STEM major status—show negligible and statistically insignificant effects. The estimated GPA effect is close to zero, while the negative point estimate for STEM major choice is small in magnitude and similarly non-significant. These findings suggest that BSP's primary impact manifests in persistence rather than academic performance conditional on enrollment or early STEM commitment. Table A6 in the appendix presents TSLS estimates using an expanded set of covariates that incorporate a broader range of pre-college characteristics—cohort year, gender, early STEM interest, and age at entry—in addition to the baseline financial and academic controls.

To assess whether BSP effects differed under the unusual conditions of the COVID-19 era, Table 2.10 separates results by entry cohort. For students entering in Fall 2020—whose first-year experience was heavily shaped by remote instruction—confidence intervals for both first- and second-year persistence contain zero and lie close to or entirely below it (e.g., [-0.253, 0.087] for completing the first year), suggesting little evidence of a meaningful program effect for this group. While the Summer 2020 iteration of BSP included both in-person and online components, which may have helped partially preserve some of the program's mechanisms, the smaller sample size (17 below, 7 above) could be one reason I have a wider confidence interval. In contrast, the results for non-COVID cohorts are far more conclusive. The confidence intervals are tight and bounded well above zero, such as [0.315, 0.824] and [0.119, 0.649] for first and second-year persistence, respectively, implying consistent and sizable program effects.

Finally, Table 2.11 presents TSLS estimates from a user-defined window of [6000, 20000] where I lower the number of observations on the right of the cutoff.[26] This fixed window mitigates concerns about sparsity in the tails of the EFC distribution and facilitates direct comparability across specifications. Given I construct these windows ad-hoc, covariates are not used in estimation under the fixed window setting. The results are consistent with prior findings: the confidence interval for first-year persistence, [0.149, 0.671], is tightly bounded

[26]Although not shown for brevity, I also check results using alternate windows of [10000,18000] , [8000, 22000], and [8000, 24000]. Effects on persitence hold under these windows.

above zero, indicating a clear and robust positive effect. For second-year persistence, the interval is wider, [-0.027, 0.518], but still suggests a positive relationship at conventional levels of significance. In contrast, the confidence interval for completed credits remains broad, spanning from -2.528 to 7.915, which reflects both statistical imprecision and potential heterogeneity in academic momentum outcomes.

Table 2.11: TSLS Estimates from Fuzzy Local Randomization RDD (Full Sample, User Window: [6000, 20000])

Outcome	**TSLS Estimate**	**95% CI**	**Obs (Below / Above)**
Complete 1st Year	0.410***	[0.149, 0.671]	87 / 33
Complete 2nd Year	0.246*	[-0.027, 0.518]	87 / 33
First-Year Completed Credits	2.693	[-2.528, 7.915]	79 / 26

Note: This table reports Two-Stage Least Squares (TSLS) estimates from fuzzy local randomization regression discontinuity designs centered at an Expected Family Contribution (EFC) cutoff of $16,000. A user-specified window of [6000, 20000] is applied across all models. Although covariates were listed, they are not incorporated in estimation when the window is set manually. Observations reflect the effective sample size (non-missing) within the specified window. Confidence intervals are based on large-sample asymptotic approximations.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Taking Stock

Taken together, the results from this and previous tables underscore that being a BSP student has a consistent and substantively meaningful effect on early college persistence, particularly for students just below the financial eligibility threshold. These effects are robust across multiple bandwidths, covariate adjustments, and cohort subsamples, with the lower bounds of the confidence intervals still indicating practically relevant improvements in retention.

By contrast, the estimates for GPA, STEM major uptake, and especially completed credits are noisier and statistically insignificant. However, this lack of precision should not be interpreted as evidence of no effect. The confidence intervals around completed credits are wide—often spanning several units—and in this context, even an increase of 3 credits would correspond to the successful completion of an additional college-level course. As such, the intervals still allow for potentially meaningful academic gains, even if I cannot confirm their statistical significance. The current data simply do not rule out important effects; rather, the estimates remain too imprecise to draw firm conclusions.

To better understand these dynamics, I conducted a descriptive comparison (not shown) of second-year spring term completed credits among students who persisted. Among persisters, the average number of earned credits was relatively similar: 11.83 credits for BSP participants and 12.60 credits for non-participants. Since students who drop out accumulate fewer credits by definition, this comparison removes much of the compositional bias. The small remaining gap in average credits underscores that BSP's primary impact is likely on keeping students enrolled rather than accelerating their academic pace.

To contextualize the TSLS results, it is also useful to compare them to the intention-to-treat (ITT) estimates in Appendix Table A7. While the two estimands differ, the findings are consistent: both approaches yield confidence intervals that suggest statistically credible, positive effects on persistence, alongside inconclusive—but not dismissible—evidence on credit accumulation. These parallels reinforce the overall conclusion that BSP exerts a strong and reliable impact on short-term college retention.

2.6 Threats to Validity and Falsification Tests

A key identifying assumption in regression discontinuity designs (RDD) is that units cannot precisely manipulate the running variable around the cutoff. If individuals can sort around the threshold, the resulting violation of the as-if random assignment undermines causal identification. To assess this, I implement a density test of the Expected Family Contribution (EFC) running variable at the \$16,000 cutoff.

Figure 2.3 displays the distribution of EFC in a neighborhood around the threshold. Visually, there is no discernible discontinuity, and formal tests (not shown for brevity) corroborate the absence of sorting. This result supports the validity of the RDD design by reinforcing the assumption that EFC is not subject to strategic manipulation. Given that EFC is a complex administrative construct derived from federal financial aid formulas and third-party inputs, this lack of manipulation is theoretically plausible and empirically verified. Thus, the density test provides additional evidence for the credibility of the local randomization framework.

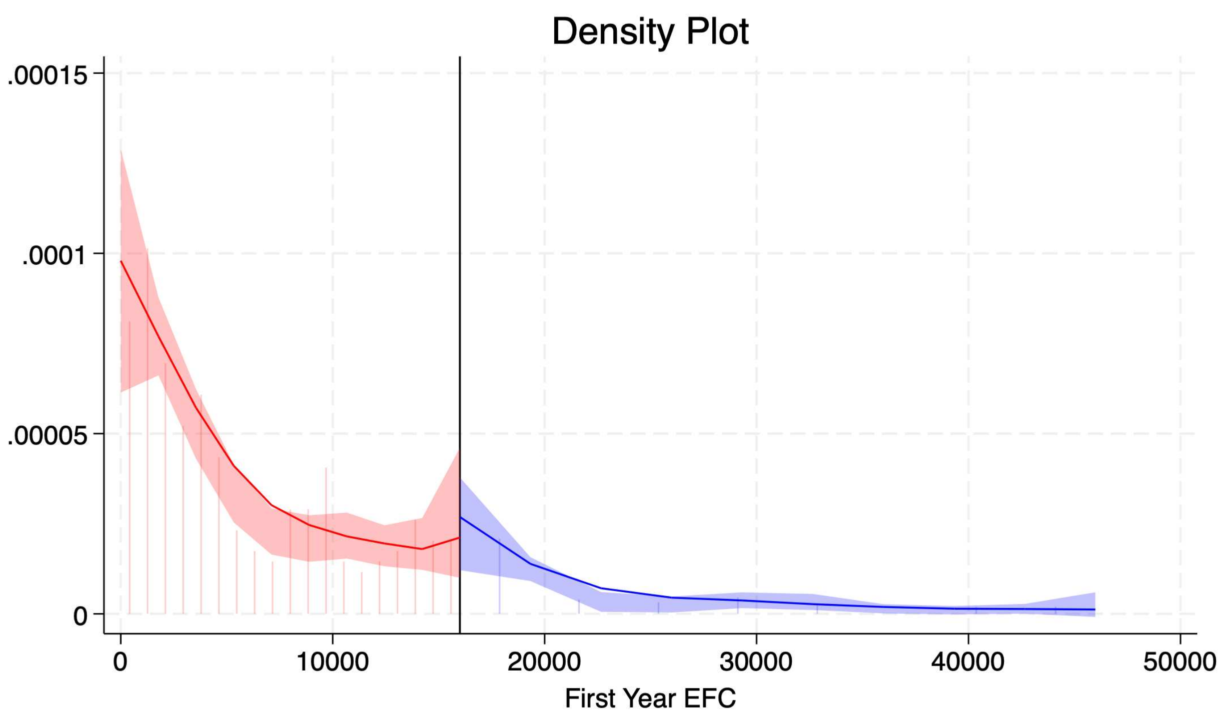


Figure 2.3: Density Test of the Running Variable (EFC)

Note: This figure shows the estimated density of the running variable (EFC) around the $16,000 threshold using a local polynomial estimator. The absence of a jump at the cutoff suggests no sorting or manipulation in response to the BSP eligibility rule.

To further assess the robustness of the identification strategy, I implement two classes of falsification tests: (1) tests using alternative cutoffs and (2) tests using pre-treatment covariates as outcomes.

First, Table 2.12 reports results from fuzzy TSLS regressions centered at placebo thresholds of $5,000 and $8,000—values at which no BSP eligibility rule applies. Across all specifications, including both second-year persistence and spring persistence outcomes, estimated effects are small in magnitude, statistically insignificant, and consistently accompanied by confidence intervals that include zero. Notably, the confidence intervals in some cases extend well beyond the logical bounds for binary outcomes (e.g., beyond $[-1, 1]$). This arises due to the unbounded nature of the TSLS estimator, especially in small samples with noisy first-stage compliance or low treatment variation, where extrapolated linear projections can fall outside the unit interval—a common feature in IV estimation with binary outcomes (Cameron & Trivedi, 2005). These placebo results reinforce confidence that the treatment effects estimated at the true $16,000 threshold are not spurious.

Table 2.12: TSLS Falsification Tests Using Alternative Cutoffs (Placebo Thresholds)

Outcome (Placebo Cutoff)	TSLS Estimate	95% CI	p-value	Obs (Below/Above)	Window [EFC]	Mean L/R	SD L/R
Second-Year Persistence ($5,000)	1.568	[-15.161, 18.296]	0.854	43 / 50	[3,000, 10,000]	0.907 / 0.880	0.294 / 0.328
Complete 2nd Year (Yr 2, $5,000)	7.378	[-59.966, 74.723]	0.830	43 / 50	[3,000, 10,000]	0.907 / 0.780	0.294 / 0.418
Second-Year Persistence ($8,000)	-0.324	[-0.776, 0.129]	0.161	25 / 34	[5,000, 11,000]	0.960 / 0.853	0.200 / 0.359
Complete 2nd Year (Yr 2, $8,000)	-0.139	[-0.736, 0.458]	0.649	25 / 34	[5,000, 11,000]	0.840 / 0.794	0.374 / 0.410

Note: Each row shows TSLS estimates under a fuzzy local randomization RDD centered at a placebo cutoff where no Bridge Scholars Program (BSP) eligibility rule applies. Confidence intervals are based on large-sample approximations. The inclusion of means and standard deviations of the outcome variable on either side of the cutoff provides additional context: when the outcome is binary and the treatment compliance rate is low or noisy, the TSLS estimator may yield extrapolated values outside the unit interval, which is a known feature of linear instrumental variable models applied to limited dependent variables (Cameron & Trivedi, 2005). Mean denotes mean values left and right of the cutoff and SD denotes standard deviation left and right of the cutoff.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Second, Table 2.13 presents placebo tests using baseline characteristics that are determined prior to BSP eligibility and should not be affected by program assignment. These include high school GPA, first-year need-based aid, merit scholarships excluding BSP awards, student loan receipt, and limited-income status as flagged by the institution. Across all specifications, I observe no statistically significant discontinuities, and the associated confidence intervals are tight and centered around zero. These findings reinforce the assumption of local balance around the cutoff and rule out endogenous sorting or manipulation of observable covariates.

Table 2.13: Falsification Tests on Baseline Characteristics

Placebo Outcome	**TSLS Estimate**	**95% CI**	**Obs (Below / Above)**
High School GPA (Rounded)	0.097	[-0.118, 0.312]	23 / 25
Merit Aid (Excl. BSP)	2329.89	[-2100.19, 6759.98]	23 / 25
Need-Based Aid (1st Year)	89.97	[-4750.27, 4930.22]	23 / 25
Loan Amount (1st Year)	-44.15	[-4426.60, 4338.31]	23 / 25
Limited-Income Status	0.052	[-0.143, 0.248]	23 / 25

Note: All regressions use fuzzy local randomization RDD centered at the $16,000 EFC cutoff with user-defined windows. These outcomes are pre-treatment and should be unaffected by BSP assignment. All estimates are statistically insignificant, and the confidence intervals consistently contain zero.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Together, these falsification tests lend credibility to the identification strategy employed in the main analysis. While small sample sizes in some placebo regressions limit precision, the general pattern of null effects across both alternative cutoffs and placebo outcomes is reassuring. No evidence suggests spurious jumps at non-policy thresholds or in baseline characteristics, bolstering the interpretation of the local randomization RDD estimates as capturing causal effects of BSP eligibility.

2.7 Possible Mechanisms

My findings suggest that the Bridge Scholars Program increases student persistence. I also cannot rule out negligible effects on total credits accumulated since the confidence intervals are not narrow enough to be considered ineffective. To better understand the processes driving these outcomes, I explore potential mechanisms that underpin the program's success. The evidence indicates that the program's unique combination of financial, social, and academic support plays a pivotal role in mitigating barriers to retention, especially for first-generation and low-income students.

2.7.1 Summer Program Effects

A key feature of the program is its eight-week summer residency, which immerses students in a structured introduction to college life. During this time, participants take introductory courses while engaging in workshops, peer activities, and community-building events. These efforts extend beyond traditional academic preparation, targeting social and cultural dimensions of the college experience.

Insights from a 2021 qualitative study conducted by college's Institutional Research office in collaboration further illuminate the relevance of these programmatic elements.[27] The study examined the experiences of first-time freshmen who left the college before their second year, focusing on barriers to persistence. The findings highlighted three predominant challenges faced by students: financial struggles, feelings of social isolation, and academic difficulties. Nearly half of the interviewed students cited financial challenges as their primary reason for leaving the college, including difficulties understanding financial aid, managing expenses, and balancing work and academics. Similarly, a significant proportion of students reported social disconnection, pointing to a lack of meaningful friendships, marginalization, and dissatisfaction with campus life. Academic challenges, while less frequently cited,

[27] Institutional Research, Planning, and Effectiveness (2021)

included insufficient preparation, poor faculty interactions, and dissatisfaction with coursework.

The Bridge program appears uniquely positioned to address these challenges. Financial literacy workshops offered during the summer residency provide participants with essential tools to navigate financial aid, manage budgets, and plan for long-term financial responsibilities. By directly targeting one of the most commonly reported barriers to persistence, these workshops reduce uncertainty and equip students with practical strategies for managing college expenses. This proactive approach likely alleviates a key source of stress for many first-generation and low-income students, thereby supporting retention.

Equally important is the program's focus on fostering community and connection. The eight-week residency creates opportunities for participants to build peer networks and establish a sense of belonging before the start of the academic year. Early friendships formed during social events, group activities, and workshops can reduce feelings of isolation and provide a support system that persists throughout the year. This approach is particularly valuable for students who may otherwise struggle to integrate into campus life, such as those from underrepresented backgrounds or geographically distant communities. By creating shared experiences, the program not only addresses social barriers but also enhances students' ability to navigate the broader college environment.

The academic component of the summer program further reinforces these benefits. Participants take introductory courses in a low-stakes setting, allowing them to adjust to college-level academics while receiving targeted support. Working closely with faculty and peers during this period can boost academic confidence and improve preparedness for the fall semester. Additionally, the workshops and advising sessions offered through the program introduce students to campus resources, such as mental health services, library facilities, and extracurricular opportunities. These resources empower students to take full advantage of the college's infrastructure, addressing the academic and personal challenges highlighted in the study.

The Bridge program's multi-faceted design aligns closely with the qualitative study's recommendations for improving retention. These recommendations emphasized the importance of financial counseling, community-building efforts, and enhanced support for off-campus students. By addressing these areas, the program provides targeted interventions at a critical juncture in students' educational trajectories. For many participants, the summer residency likely serves as a bridge not only to academic success but also to a sense of belonging that sustains them throughout their college journey.

2.7.2 Scholarship Effect

A key feature of BSP is its $2,500 annual scholarship, amounting to up to $10,000 over four years. This award is distributed only after students successfully complete the summer program and is not captured in the baseline financial aid or merit-based scholarship values used in the local randomization RDD models. While I show BSP participants and non-participants near the EFC cutoff are balanced on reported financial aid, this additional $2,500 allocation could still plausibly contribute to the observed effects on college persistence.

To better contextualize whether the observed BSP effects might be driven, at least in part, by this supplemental financial aid, I examine the impact of three other scholarships of similar monetary value. These analyses use causal designs—including fuzzy regression discontinuity (RDD), regression kink design (RKD), and difference-in-differences (DiD)—applied to the full student population, excluding the 410 BSP applicants. While these samples differ demographically and academically, they nonetheless offer useful suggestive evidence about whether financial aid of this magnitude, offered independently of structured programming, can influence student outcomes. These analyses are drawn from an ongoing internal working paper (Shakya, 2025).[28]

I begin with a merit based scholarship offered to students based on their high school GPA at the time of application. Using a fuzzy RDD centered at the 3.9 GPA threshold, I find

[28] A brief summary of the paper is discussed in Appendix 3.6

large and statistically significant effects of this scholarship on second-year persistence and credit accumulation among pre-COVID cohorts—magnitudes comparable to BSP estimates. However, these effects do not hold among post-COVID students. One possible explanation is that pandemic-related changes (e.g., remote learning, inconsistent grading practices, and grade inflation) may have weakened the informational value of GPA as a good signaling mechanism.

Next, I apply a regression kink design to study need-based institutional aid that increases at an EFC of $6,000. While this design exploits variation in aid allocation slopes rather than discontinuities, the results show no robust or precise effects on GPA or persistence. Still, it would be premature to claim null effects—the upper and lower bounds of the confidence intervals remain wide enough to include substantively meaningful outcomes. In other words, these results do not rule out the possibility of positive effects.

Lastly, I study a high-school specific award (hereby called Alliance Award), which targets students from ten designated partner high schools across the state. These schools were selected based on a combination of high rates of free/reduced lunch eligibility, high first-generation college-going rates, and regional representativeness across the state. Students from Alliance schools are among the most structurally disadvantaged and academically underprepared at the college. I use a DiD framework to evaluate a 2016 policy change that raised the Alliance Award from $2,500 to $4,000. The findings suggest modest improvements in GPA and persistence for the earliest cohorts following the increase, but these effects attenuate over time.

Taken together, these analyses suggest that modest scholarships may help nudge student outcomes in the right direction—but in no case do the effects match the magnitude or durability of those observed for BSP. The BSP scholarship does not operate in isolation: it is embedded within a broader framework of mentoring, academic scaffolding, and social integration. Unfortunately, the current design does not permit a clean test of whether the scholarship alone drives these outcomes. Ideally, one would compare BSP students who receive the scholarship to those who do not—but such variation does not currently exist. A promising future direction would involve assigning the $2,500 scholarship to non-selected BSP applicants (or others from a similar pool) in cases where capacity or funding limits preclude full program enrollment. Doing so would provide a clearer test of whether financial support alone can replicate BSP's impacts.

Importantly, while my scholarship analyses do not conclusively show strong and persistent effects, they also do not provide precise null results and confidence intervals often span meaningful values. This suggests that scholarships may be influencing persistence and performance—but not in a manner strong or consistent enough to explain BSP effects on their own. In short, scholarships may be one part of the story, but they are unlikely to be the whole story.

2.8 Conclusion

This paper evaluates the impact of the Bridge Scholars Program (BSP), a summer intervention designed to support first-generation and low-income students, using a fuzzy regression discontinuity design under a local randomization framework. The primary estimand throughout is the treatment effect on the treated, estimated via Two-Stage Least Squares (TSLS), supplemented by intention-to-treat (ITT) results in the appendix for interpretative robustness.

The findings offer strong evidence that BSP eligibility substantially improves early college persistence. In every specification where the local window satisfies covariate balance, confidence intervals for first-year and second-year persistence exclude zero and lie entirely in the positive range—even under conservative window definitions or alternative covariate sets. At their lower bounds, these intervals still suggest practically meaningful gains: for example, a five percentage point increase in second-year persistence is the most conservative interpretation of the main estimate. This robustness, rather than the magnitude of point estimates alone, serves as the most compelling evidence for BSP's effectiveness.

By contrast, the program's impact on credit accumulation, GPA, and early STEM take-up is more ambiguous. Confidence intervals for these outcomes are wider and span zero. However, even here, the intervals often include effect sizes that are educationally meaningful. For instance, the lower bound of the completed credits estimate corresponds to nearly one full course in a semester—a nontrivial margin of uncertainty. So, in this case, I cannot strongly argue that the program has no effects on course completion.

While I show that students in the local randomization inference approach did not have different merit based and need based aid awards, BSP includes a $2,500 annual scholarship which could be driving the positve effects on persistece. Comparative analysis of other small-scale financial aid programs suggests that this additional aid alone is unlikely to fully explain BSP's effects. These other programs show at best modest and inconsistent effects on persistence and GPA, reinforcing the view that BSP's bundled supports—structured mentorship, summer programming, and peer networks—play a central causal role.

From a methodological standpoint, this study contributes to a cery limited body of literature using local randomization RDD by applying both TSLS and ITT estimators under Fisherian and asymptotic frameworks. While the Fisherian approach prioritizes randomization-based inference and hypothesis testing, the use of asymptotic confidence intervals enables interpretability of estimated ranges and their relevance for policy. The convergence of findings across both frameworks enhances the credibility of the results.

Looking ahead, future work should aim to experimentally disentangle BSP's bundled components—especially by varying the provision of the scholarship independently of the summer program. One practical avenue could involve randomizing aid among non-selected BSP applicants, or extending similar supports to other underrepresented students under capacity constraints. Longitudinal follow-up is also needed to determine whether short-term persistence gains lead to higher graduation rates and improved post-college outcomes.

In sum, this paper shows that the Bridge Scholars Program significantly improves short-run persistence among students near the EFC eligibility threshold. These gains are robust to specification, bandwidth choice, and inferential approach. By addressing financial, academic, and social barriers simultaneously, the program not only improves short-term retention but also lays a foundation for long-term achievement. These findings reinforce the importance of multi-faceted strategies that extend beyond financial aid, emphasizing the need to foster belonging, preparedness, and resilience among underrepresented student populations. As higher education institutions strive to close equity gaps and promote upward mobility, programs like BSP offer a promising model for cultivating inclusive and supportive campus environments. While the program's early results are encouraging, it is important to acknowledge its relatively limited scale and nascent stage of implementation. Even under conservative assumptions, we find evidence of a positive impact on student persistence; however, effects on longer-term outcomes remain imprecise. Future research should continue to rigorously evaluate the program's effectiveness, both to assess its sustained impact and to inform ongoing efforts to refine and enhance its structure.

Chapter 3

The Pandemic Learning Shock That Wasn't: STEM Major Choice Resilience During COVID-19

This study investigates whether the COVID-19-induced shift to remote learning impacted undergraduate students' decisions to remain in or switch into STEM majors. Using longitudinal administrative data from a large U.S. public university (cohorts 2013–2019) and a staggered difference-in-differences design, I find no meaningful effects of remote instruction on whether college students remain as STEM majors. Robustness checks, including nearest-neighbor matching, confirm negligible average effects suggestive that even when assuming the true estimates are at the upper end of our estimated confidence intervals, the effect is negligible.. Overall, the evidence suggests that students' field-of-study choices were largely resilient to the instructional format change, at least in the short run.

3.1 Introduction

The COVID-19 pandemic dramatically disrupted higher education by forcing an unprecedented shift from traditional in-person instruction to remote learning. This rapid transition raised critical questions regarding its implications for student engagement, learning outcomes, and retention, particularly in academically demanding fields such as Science, Technology, Engineering, and Mathematics (STEM). The urgency of understanding these impacts is heightened by persistent under-representation and performance disparities among African American and Hispanic/Latino students in STEM disciplines and how the pandemic might have affected different college students differently. (Arcidiacono et al., 2016; Barber et al., 2021; Griffith, 2010). While existing research has shown that the pandemic had negative effects on student outcomes at the high school level (see Ardington et al. (2021), Azevedo et al. (2021), Bacher-Hicks et al. (2021), and Jack et al. (2023)), could the pandemic have also shifted students' interest in pursuing STEM majors?

A substantial body of research highlights the role of academic preparation, institutional environment, and financial aid in influencing STEM persistence (Bettinger & Long, 2009; Griffith, 2010). Institutional characteristics, including available resources and support structures, play a critical role in shaping the experiences and retention of minority students in STEM fields (Bettinger & Long, 2009; Griffith, 2010). Historical evidence suggests that economic downturns and policy shifts can also significantly alter enrollment patterns and academic decisions, thereby exposing the vulnerability of student trajectories to external shocks (F. Y. Ersoy, 2020; Long, 2014). In the unique context of the COVID-19 pandemic, early evidence indicates that the sudden shift to remote learning may have disproportionately disadvantaged certain student populations by exacerbating pre-existing inequalities (Aucejo et al., 2020; Bird et al., 2022; Rodriguez-Planas, 2022).

Existing empirical research suggests that the pandemic's effects on academic performance in college were relatively limited. For instance, Mostafa et al. (2023) analyze the flexible grading policies adopted by a public university in North Carolina during the pandemic and find that while usage varied by course type and student background, the policy did not produce widespread academic distortions. Similarly, Bird et al. (2022) estimate modest negative effects (3–6%) on course completion in a statewide community college system but find little evidence of persistent academic consequences from the shift to online learning. Taken together, these findings suggest that academic outcomes remained relatively stable– mainly due to the flexible grading policies most students took advantage of. This being said, broader disruptions—such as changes in perceived major value, labor market uncertainty, or remote instruction experiences—could still have influenced students' major switching behavior in more subtle but important ways.

The potential effects of COVID (and the associated remote instruction shock) on STEM major retention are theoretically ambiguous. I consider two competing hypotheses: (1) remote learning may deter students from staying in or switching into STEM majors due to a loss of in-person laboratory access and academic support; or (2) it may instead encourage students to persist in or even switch into STEM, thanks to a surge of interest in health-related fields during the pandemic and perhaps more flexible online coursework. Recent studies (Pertold-Gebicka, 2024; Rodríguez-Planas, 2022) have shown how students took advantage of the lenient grading policies colleges implemented during the pandemic. The first hypothesis emphasizes that the abrupt transition to online learning removed key experiential and social aspects of STEM education. Without hands-on lab work and face-to-face interaction with instructors and peers, students might find it harder to stay motivated and succeed in STEM classes, potentially prompting them to switch out to other disciplines. This concern is especially true in lab-intensive majors (such as engineering and biology), where in-person experimentation and close mentorship are central to the learning experience (Chen et al., 2021). The second hypothesis highlights potential benefits of the online format. Remote learning can offer greater flexibility and lower pressure—for example, through recorded lectures, self-paced study, or open-book exams—which might make challenging STEM coursework feel more manageable. Additionally, the COVID-19 crisis dramatically raised public awareness of health and science, possibly inspiring more students to pursue studies in biology and related fields. While students' decisions about college majors are shaped by a wide range of academic, personal, and structural factors (F. Ersoy & Speer, 2025), it is plausible that the COVID-19 pandemic—through its disruption of instruction, campus life, and broader social conditions—altered the trajectory of major choices for students enrolled during this period.

This paper examines whether the pandemic-induced shift to remote learning had any differential influence on students pursuing STEM majors. In addition to examining STEM, I also assess whether the pandemic increased the number of students opting into biology and health-related fields. The core focus is on whether COVID-19—through both its educational disruptions and broader societal impacts—altered students' major choice decisions. The analysis uses administrative data from a large public land-grant university in the United States. The dataset includes semester-level information on students' majors, academic performance, and financial aid, and spans undergraduate cohorts entering between Fall 2013 and Fall 2019.

To estimate causal effects, I implement a staggered Difference-in-Differences (DiD) design that exploits variation in when different cohorts were exposed to remote learning. At this university—as with many U.S. colleges—instruction shifted fully online at the end of March 2020 (mid-Spring 2020), with many courses remaining remote or hybrid through Fall 2020. By Fall 2021, nearly 95% of classes had returned to in-person instruction. The design leverages variation in when cohorts hit their "pandemic semester" to evaluate whether the switch to remote learning influenced major switching behavior. For example, students entering in Fall 2019 experienced remote instruction in their third semester, while earlier cohorts encountered it later (if at all) in their academic careers. This setup enables a comparison of within-student changes across cohorts, attributing deviations from historical switching patterns to the remote learning shock. I only look at students that continued their college studies up until the end of their eight semester and thus, this analysis only looks at students who did not drop out. The effects of the pandemic on college enrollment have been well documented and are not within the scope of this paper (Koopmann et al., 2024; Olmedo-Cifuentes & Martínez-León, 2022; Ye et al., 2022).

I find that remote instruction had minimal overall influence on STEM major retention and switching behavior. Across a range of model specifications, there is no statistically significant effect of the remote-learning transition on students' likelihood of staying in a STEM major or switching into STEM. Quantitatively, the estimated impacts are centered near zero. The magnitude of this effect is, at most, on the order of 0.015–0.025 standard deviations, indicating a negligible change in practical terms if we assume that the true estimates are on upper or lower ends of the confidence intervals. I also document a brief drop in the rate at which students switched into STEM or biology majors in the immediate aftermath of the shift (particularly during Fall 2020). However, this dip appears to be temporary—switching rates rebound to baseline within one to two semesters. In short, despite the unprecedented disruption to instructional delivery and campus life, students' academic pathways in STEM remained largely unchanged.

This study contributes to the economics of education literature by providing new empirical evidence on how a large, exogenous shock to instructional delivery affected field-of-study decisions among college students. While anecdotal accounts and media narratives have speculated that COVID-19 could dramatically reshape students' academic trajectories—either by deterring persistence in STEM due to remote lab closures or by inspiring greater interest in health-related fields—the findings suggest otherwise.[29] The pandemic did not produce large shifts in STEM participation nor did it significantly increase transitions into biology or pre-medical tracks. This relative invariance may reflect the resilience of student preferences and the strength of long-term educational commitments that were formed prior to the

[29]For instance, reports have documented how the pandemic disrupted hands-on lab courses, potentially deterring students from persisting in STEM majors (Prudencio et al., 2024). Additionally, shifts in academic trajectories towards more stable and financially secure career paths were observed, such as in business and supply chain management programs (Levanon, 2023; P. Page, 2023). Conversely, heightened interest in healthcare careers was noted among some students, aligning with the increased visibility of healthcare workers during the pandemic (Law, 2021; Xie, 2021).

pandemic. It also implies that neither the remote learning modality nor heightened visibility of healthcare crises during the pandemic was sufficient, on its own, to drive large-scale changes in the academic trajectory of students who were already enrolled in college before the COVID-19 pandemic started.

3.2 Data and Sample Selection

This study leverages rich administrative data from a large public land-grant university in the United States, covering undergraduate cohorts who entered between Fall 2013 and Fall 2019. The dataset provides detailed semester-by-semester records of students' academic outcomes, demographic characteristics, and financial aid status.

To construct a consistent and balanced panel for analysis, I apply several restrictions. First, I focus exclusively on first-time, full-time undergraduate students who started college in the fall term. Transfer students are excluded due to heterogeneity in prior academic preparation and credit transferability. Second, I retain only domestic students aged 17–19 at entry, and also omit international students because their experiences may be systematically different. Finally, I limit the sample to students who remained enrolled through at least their eight academic semester (i.e., through the end of their fourth year), ensuring that the observed outcomes reflect genuine major persistence or switching rather than attrition. In the robustness checks presented later in this paper, I demonstrate that these restrictions—particularly the exclusion of dropouts—do not bias comparisons across cohorts. After applying these criteria, the final analytic sample consists of 20,108 unique students. I then restructure the dataset into a panel format to facilitate longitudinal analysis. College major data is collected both at the start of the semester (census) and at the end of the term.

Table 3.1: Descriptive Statistics of Fixed Student Characteristics by Entry Cohort

Variable	**FA13**	**FA14**	**FA15**	**FA16**	**FA17**	**FA18**	**FA19**
High School GPA	3.647	3.647	3.636	3.675	3.665	3.690	3.738
	(0.428)	(0.432)	(0.458)	(0.457)	(0.467)	(0.460)	(0.450)
Merit-Based Aid ($)	2,460	2,581	3,000	3,687	3,783	3,741	3,612
	(4,887)	(4,690)	(5,181)	(6,017)	(6,585)	(6,260)	(5,762)
Need-Based Aid ($)	2,432	2,291	2,432	2,761	2,632	2,516	2,987
	(4,828)	(4,914)	(5,107)	(5,328)	(5,378)	(5,403)	(5,912)
First-Gen (1 = Yes)	0.223	0.204	0.241	0.212	0.196	0.197	0.203
	(0.417)	(0.403)	(0.428)	(0.409)	(0.397)	(0.398)	(0.402)
Limited-Income (1 = Yes)	0.229	0.216	0.213	0.259	0.237	0.211	0.243
	(0.420)	(0.412)	(0.410)	(0.438)	(0.425)	(0.408)	(0.429)
Female (1 = Yes)	0.552	0.548	0.544	0.569	0.547	0.554	0.571
	(0.497)	(0.498)	(0.498)	(0.495)	(0.498)	(0.497)	(0.495)

Note: Table reports means with standard deviations in parentheses. All variables are measured at the point of college entry (i.e., wave = semester of matriculation). Financial aid variables reflect total merit- and need-based aid awarded in the first academic year. "Female" is coded as 1 for female students, 0 for male. Institutional data for gender are only available in binary categories (Male/Female).
$N = 20{,}108$

Tables 3.1 and 3.2 present descriptive statistics for the main covariates and outcomes used in the analysis. Table 3.1 summarizes fixed student characteristics by entry cohort.

Average weighted high school GPA[30] is approximately 3.67 across cohorts, with consistent trends in financial aid: students receive an average of $3,281 in merit-based and $2,584 in need-based financial aid during their first year across cohorts. Roughly 21% of students are identified as first-generation, and 23% come from limited-income backgrounds. Gender is recorded as binary in institutional data, with just over half of students identifying as female.

Table 3.2 reports means and standard deviations for semester-varying outcomes, based on the student's time in college (e.g., semester 1 = first term enrolled). For my analysis, I use the major recorded at the end of the term. The primary reason for using the end-of-term major rather than the census major is to avoid the potential bias that could arise from using the initial major declaration. Since students have approximately three months during their first semester to make more informed decisions about their major, the end-of-term major is likely a more accurate representation of their academic commitment.

Across semesters 1 through 5, STEM major participation remains stable at approximately 41–42%, while the proportion of students in biology-related fields gradually declines from 21% to 17%. These include majors in the following departments: "Biology", "Biochemistry", "Molecular Biology", "Biomedical Sciences", "Microbiology, Immunology, and Pathology", "Radiological Health Sciences", and "Health and Exercise Sciences". I hereby refer to these groups of majors as Bio-related majors. Switching behavior is most concentrated in the second and third semesters, with about 2.7% of students switching into STEM and 3.3% switching out of STEM in semester 2. Switching into biology-related majors occurs at much lower rates, with fewer than 1% of students doing so in any given term. Overall departmental switching is most frequent in the early semesters, affecting 13.9% of students in semester 2, then tapering off in later terms.

[30] A weighted GPA is scaled based on the difficulty level of a course, with more challenging courses being worth more. The scale for calculating weighted GPA can exceed 4.0

Table 3.2: Descriptive Statistics of Semester-Varying Outcomes by Student Semester

Variable	Semester 1	Semester 2	Semester 3	Semester 4	Semester 5
STEM Major (1 = Yes)	0.417	0.411	0.419	0.419	0.419
	(0.493)	(0.492)	(0.493)	(0.493)	(0.493)
Bio-Related Major (1 = Yes)	0.209	0.196	0.184	0.173	0.167
	(0.406)	(0.397)	(0.387)	(0.378)	(0.373)
Switch to Bio-Related Majors (1 = Yes)	0.000	0.009	0.009	0.005	0.005
	(0.000)	(0.096)	(0.093)	(0.073)	(0.067)
Switch into STEM (1 = Yes)	0.000	0.027	0.034	0.021	0.016
	(0.000)	(0.162)	(0.180)	(0.143)	(0.127)
Switch out of STEM (1 = Yes)	0.000	0.033	0.026	0.021	0.016
	(0.000)	(0.178)	(0.160)	(0.143)	(0.125)
Switch Departments (1 = Yes)	0.000	0.139	0.109	0.080	0.068
	(0.000)	(0.346)	(0.312)	(0.272)	(0.252)

Note: Table reports means with standard deviations in parentheses. "Semester" refers to a student's term in college (Semester 1 = first enrolled term, Semester 2 = second term, etc.). All switching indicators are coded as 1 only in the semester when the change in major occurred, and 0 otherwise. Observations are at the student-semester level. Showing only upto the 5th semester for brevity.
N = 20,108

Table 3.3 presents descriptive statistics on major enrollment and switching behavior by cohort, based on the reshaped panel dataset where each row represents a student-semester observation. This structure increases the total number of observations to 160,864, as each of the 20,108 unique students contributes up to eight semester-level records. The distribution of student-semester observations is consistent across cohorts.

Switching into STEM or biology majors is relatively uncommon: only 2,167 students (roughly 10.8% of total STEM participants) switched into STEM from a non-STEM field, and just 1,132 students (about 4.1% of biology participants) switched into biology from another field. The incidence of switching is highest in earlier cohorts (e.g., FA13–FA15), with modest declines in more recent ones. These patterns suggest that while broad departmental switching is fairly common across the college experience, targeted moves into STEM—and particularly into biology—are more selective. These cohort-level trends are broadly consistent with national data from postsecondary institutions (National Center for Education Statistics (NCES), 2017).

Table 3.3: STEM and Bio-Related Participation and Switching by Entry Cohort (Panel Format)

Variable	**FA13**	**FA14**	**FA15**	**FA16**	**FA17**	**FA18**	**FA19**
STEM Majors	8,696	8,779	9,390	9,327	9,376	10,510	10,828
Bio-Related Majors	3,742	3,610	3,723	3,894	3,797	4,206	4,448
Switched into STEM	339	350	316	297	300	290	275
Switched out of STEM	322	341	321	304	343	337	279
Switched into Bio-Related	182	170	168	161	149	166	136
Student-Semester Observations	22,280	21,528	23,040	23,032	22,808	24,216	23,960

Note: Table reports student-semester level counts by cohort of entry. Each count reflects the total number of student-semester observations that meet the specified criterion. For example, "Switched into STEM" captures the number of student-semesters in which a student first switched into a STEM field. The dataset is structured in panel format, and students can appear in up to eight semesters. While the total number of unique students is 20,108, the full dataset includes $20{,}108 \times 8 = 160{,}864$ observations.

The goal of the study is to evaluate whether the onset of COVID and the disruption it created affected college major choices of affected students. For the initial preliminary analysis, I first employ a simple Ordinary Least Squares (OLS) regression.

3.3 Preliminary Analysis: OLS

In order to assess how the shift to remote instruction may have influenced major persistence, I estimate student-level fixed-effects regressions of the following form:

$$\text{Outcome}_{it} = \alpha_i + \beta_1 \text{D}_{it} + \beta_2 \text{GPA}_{it} + \beta_3 \text{MeritAid}_{it} + \beta_4 \text{NeedAid}_{it} + \sum_{s=2}^{8} \gamma_s \text{Semester}_{s,it} + \varepsilon_{it} \quad (3.1)$$

where Outcome_{it} is a binary indicator for whether student i is enrolled in either a STEM or biology-related major in semester t. The key explanatory variable, D_{it}, is an indicator that equals 1 starting in Fall 2020 for students who were still enrolled at that time, capturing exposure to remote instruction during the COVID-19 pandemic. Once treated, the treatment does not go back to zero.

The variable $\text{Semester}_{s,it}$ denotes the student's semester-in-college, i.e., their academic tenure (e.g., 1st semester = 1, 2nd = 2, etc.), with $s = 1$ as the omitted category. These semester indicators control for generic patterns of major switching or persistence over academic progression.

Student fixed effects α_i absorb all time-invariant individual characteristics, such as baseline academic preparedness, cohort wave (year of entry), or demographic background. The regression also includes time-varying covariates: end-of-term GPA, merit-based aid, and need-based aid, all measured at the semester level.

Table 3.4: OLS Fixed Effects Estimates for Major Enrollment and Switching

	STEM Major Enrollment		Biology Major Enrollment		Departmental Change	STEM Change
Variable	**No Controls**	**With Controls**	**No Controls**	**With Controls**	**With Controls**	**With Controls**
COVID Impact	-0.0013	-0.0024	-0.0105***	0.0041**	0.0001349	0.000073
	(0.0020)	(0.0020)	(0.0013)	(0.0015)	(0.0002019)	(0.000157)
GPA	X	-0.0279***	X	-0.0344***	-0.0004919**	-0.0003303***
		(0.0009)		(0.0013)	(0.0001637)	(0.0001273)
Merit-Based Aid	X	0.0121***	X	0.0010	-0.0000724	-0.000144
		(0.0015)		(0.0011)	(0.0001542)	(0.0001199)
Need-Based Aid	X	-0.0023	X	-0.0038***	-0.0002119	-0.0002492**
		(0.0014)		(0.0011)	(0.000147)	(0.0001143)
Semester Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Student Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	160,864	157,130	160,864	160,554	160,554	160,554
Number of Students	20,108	20,108	20,108	20,108	20,108	20,108
Within R^2	0.0002	0.0085	0.0005	0.0085	0.0006	0.0002

Note: Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Columns (1)–(2) present fixed-effects estimates for STEM major enrollment; Columns (3)–(4) report estimates for Biology major enrollment. Column (5) presents estimates for departmental change (probability of starting in a different department and ending in a different department within the same semester). Column (6) presents estimates for starting as a STEM major and ending non-STEM that very semester and vice-versa. All models include student and semester fixed effects.

The estimates reveal mixed patterns. For STEM major enrollment, the COVID coefficient is small and statistically insignificant, suggesting that remote instruction had little detectable average effect on STEM persistence. In contrast, biology major enrollment exhibits a slight negative association in the unadjusted model, which flips to a small positive and statistically significant effect once controls are included. While this suggests a marginal increase in biology enrollment following the shift to remote learning, the magnitude is modest (considering the mean values of percentage of students majoring in STEM and Biology-related majors). Likewise, for these groups of students, the pandemic semesters seems to have had no effect on their decision to switch majors that particular semester (Spring 2020).

Although the fixed-effects OLS estimates for biology-related majors are statistically significant, their interpretation warrants caution. The specification in equation (3.1) captures the average within-student association between remote instruction and major persistence, controlling for time-varying academic and financial variables as well as semester-in-college fixed effects. However, this approach is limited in several important ways.

First, the model assumes strict exogeneity, meaning that no unobserved time-varying shocks simultaneously affect both treatment status and outcomes. This assumption is particularly problematic in the context of the COVID-19 pandemic, which introduced numerous concurrent disruptions—ranging from financial instability and health crises to shifts in career preferences—that plausibly correlate with academic trajectories.

Second, the OLS model imposes a homogeneous treatment effect and does not account for dynamic responses—such as whether students are more likely to switch majors immediately following remote exposure or only after several semesters. Nor does it allow for the possibility that the effect of remote instruction differs across cohorts (e.g., Fall 2018 vs. Fall 2019), a form of treatment effect heterogeneity that may be central to understanding the policy's impact.

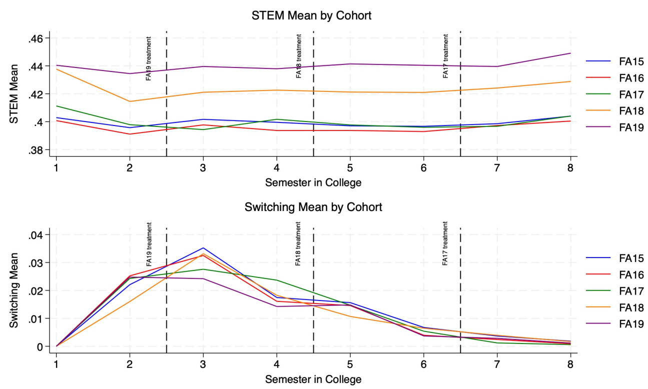
Third, while student fixed effects absorb time-invariant characteristics—including entry cohort (wave)—the OLS model implicitly treats earlier cohorts as valid counterfactuals for later ones. This may be violated if secular trends or institutional changes were already influencing major selection independently of remote instruction.

To address these limitations, the next section introduces a more credible empirical strategy: a staggered Difference-in-Differences (DiD) design based on the framework of de Chaisemartin and D'Haultfœuille (2020). This approach accommodates variation in treatment timing and allows for cohort-specific causal effects, thus overcoming key identification challenges associated with standard two-way fixed-effects DiD estimators when treatment effects are heterogeneous or dynamic.

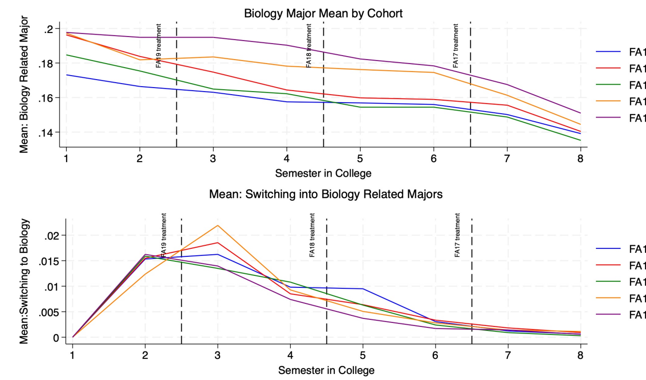
3.4 Methodology

To credibly estimate the impact of COVID-19-induced remote learning on students' decisions to persist in STEM majors, I implement a staggered Difference-in-Differences (DiD) methodology. Traditional DiD approaches compare a single treatment group and control group before and after treatment. However, in the setting, different student cohorts experienced the onset of remote instruction at varying points in their academic trajectories, leading to a staggered adoption of the treatment. Specifically, treatment onset occurred in Fall 2020 for all cohorts, but the academic semester during which students were first exposed to remote learning varied systematically by cohort entry group.

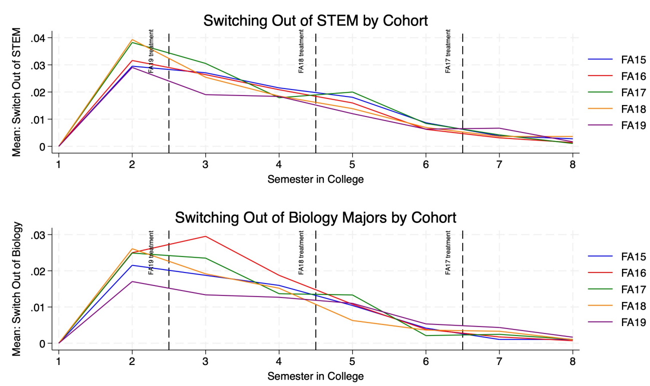
To illustrate the staggered nature of the treatment, I present figures that illustrate how different cohorts experienced the pandemic induced learning effects during different semesters of college.



(a) STEM Majors



(b) Biology/Health Majors



(c) Switching Out

Figure 3.1: Raw Means across Semesters (STEM, Biology/Health, Switching Out)

Note: The three panels display aggregate means by cohorts across semesters in college. Panel (a) presents means for STEM majors, panel (b) for Biology/Health-related majors, and panel (c) for students switching out of their original major. Fall 2015 and Fall 2016 cohorts were not affected by COVID-19 until late in their 8th semester, serving as quasi-control groups for comparison.

Figure 3.1 a) plots, by entering cohort, the proportion of students who are enrolled in a STEM major (upper panel) and the proportion who transition into a STEM major from a non-STEM field (lower panel) during each semester of their undergraduate careers. Both outcomes are binary indicators, and the vertical dashed lines mark the first semester in which a given cohort experienced COVID-19–related disruptions. Figure 3.1 b) presents the analogous statistics for Biology/Health-related majors. Figure 3.1 c) shows the proportion of students switching out of either STEM or bio-related majors. Across all six outcome measures, the trajectories are remarkably similar across cohorts: cohort-specific curves are nearly parallel, suggesting that pre-COVID academic trajectories are comparable. One interesting pattern to note is that for later cohorts (for example, take FA19 cohorts), the rate of both switching in and switching out of STEM and bio-related majors is lower compared to other cohorts.

Because the identification strategy compares students who are at the same point in their academic "tenure" rather than in the same calendar year, semester-in-college is the relevant time dimension. I therefore assess the effect of the pandemic by contrasting behavior in a given semester (hereby referred as time-in-college) for cohorts exposed to COVID-19 with behavior in that same semester for earlier cohorts that were unexposed.

The empirical setting for this study involves staggered treatment timing: students across different entry cohorts experience remote instruction at different points in their academic careers. For instance, students who entered in Fall 2019 experienced the shift to remote learning during their third semester, while students who entered in Fall 2017 encountered the same shift in their sixth semester. This design allows for the estimation of not only average treatment effects, but also dynamic effects that vary with the timing of exposure.

To begin, I illustrate the conventional econometric framework used to capture dynamic treatment effects in such settings:

$$Y_{igt} = \alpha_i + \lambda_t + \sum_{k \neq 0} \beta_k \mathbf{1}(t - E_g = k) + X'_{igt}\beta + \varepsilon_{igt} \tag{3.2}$$

Here, Y_{igt} is the outcome of interest (e.g., STEM enrollment) for student i from cohort group g in semester term t; α_i denotes student fixed effects which also captures cohort (entry-wave) effects, and λ_t are time in college fixed effects. The terms β_k capture dynamic effects relative to the semester a student was first treated with $(t - E_g = k)$ an indicator for being k years from the treatment starting, and X_{igt} includes time-varying covariates such as GPA and financial aid. Although widely used, this two-way fixed effects (TWFE) formulation has been shown to suffer from serious limitations in staggered treatment contexts when treatment effects are heterogeneous across groups or over time. Recent literature demonstrates that the TWFE estimator in such designs can produce biased results because it relies on inappropriate comparisons—specifically, using already-treated units as controls for newly treated ones (de Chaisemartin & D'Haultfœuille, 2020; Goodman-Bacon, 2021b). To address this, I follow the approach proposed by de Chaisemartin and D'Haultfœuille (2020) and Callaway and Sant'Anna (2021), by constructing group-time average treatment effects using only valid comparisons between treated units and units that are not yet treated at the time of exposure.

Formally, for a group g treated in period E_g, the causal parameter of interest is the group-time average treatment effect:

$$\mathrm{ATT}_{g,t} = \mathbb{E}[Y_{it}(1) - Y_{it}(0) \mid G_i = g, T_i = t] \tag{3.3}$$

This estimand measures the difference in potential outcomes between treated and untreated conditions for students in group g at time t. Identification relies on a conditional parallel trends assumption:

$$\mathbb{E}[Y_{it}(0) - Y_{i,t-1}(0) \mid G_i = g, X_i] = \mathbb{E}[Y_{it}(0) - Y_{i,t-1}(0) \mid G_i = c, X_i] \tag{3.4}$$

for all untreated groups c at time t, conditional on observed covariates X_i. This assumption permits group-specific, time-varying shocks, as long as they are uncorrelated with treatment conditional on observables.

Unlike the TWFE model, which includes covariates linearly in a regression, this estimator incorporates them nonparametrically using propensity score weighting or matching techniques. I select the covariates used in the model based on both theoretical relevance—such as academic preparation and financial aid eligibility—and empirical predictive power. To inform covariate selection, I run a Random Forest classifier on pre-treatment outcomes (Appendix 3.6). This exploratory exercise confirms that high school GPA, SAT scores, high school rank, and early college GPA trajectories are among the strongest predictors of major choice. While machine learning aids in selecting relevant covariates, the causal identification strategy rests entirely on the conditional parallel trends assumption and valid control group construction.

The $\text{ATT}_{g,t}$ estimates are aggregated across groups and event times to produce a dynamic event-study visualization of treatment effects. I estimate these effects for up to three semesters following exposure to remote instruction and conduct placebo tests in the two semesters prior. Standard errors are clustered at the cohort-group level to account for within-group serial correlation.

While the staggered DiD design and associated event-study framework offer a rigorous methodology for identifying the causal impact of COVID-19-induced remote instruction on STEM persistence, several critical considerations related to internal validity must be explicitly addressed. The validity of the staggered DiD estimates fundamentally relies on the parallel trends assumption, stipulating that in the absence of treatment, different cohorts would have experienced comparable trajectories in STEM retention over semesters in college. A violation of this assumption may occur if external factors unrelated to COVID-19 systematically influenced certain cohorts at specific semesters. For instance, institutional policy changes, curricular reforms within STEM disciplines, shifts in academic advising practices, or broader economic trends that altered perceptions of STEM careers might differentially impact certain cohorts precisely when remote instruction began. Such external shocks, if correlated with the timing of remote instruction, could bias the treatment effect estimates, complicating causal interpretation.

Despite these untestable methodological challenges, the staggered DiD event-study framework remains the most appropriate strategy in this context. The sudden, externally imposed transition to remote instruction due to the COVID-19 pandemic provides a credible source of exogenous variation and allows us to get closer to a causal impact compared to alternative methodologies. For robustness, I also employ the nearest neighbor matching method to test if the estimates from the staggered DiD are wildly different from that obtained from nearest neighbor matching. I now discuss the results of the study.

3.5 Main Results

I begin by presenting the results of a staggered Difference-in-Differences (DiD) event study examining how the shift to remote learning affected students' major choices. the identification strategy exploits variation in the timing of COVID-19 exposure across entry cohorts. Specifically, I compare the Fall 2019 freshman cohort—who encountered the pandemic-induced remote instruction in their *third* semester (Fall 2020)—against earlier cohorts (e.g., Fall 2017 and Fall 2016 entrants) for whom the remote-learning disruption occurred later in their college tenure. Similarly, I contrast the Fall 2018 entrants (first exposed to remote learning in their *fifth* semester, also Fall 2020) with cohorts entering in Fall 2014–2017, none of whom experienced remote instruction by their fifth semester. The identifying assumption is that in the absence of the pandemic, students across cohorts would exhibit similar semester-by-semester patterns of major switching and persistence; any deviation in a treated cohort's behavior (relative to the control cohorts) at the time of the COVID shock can thus be attributed to the impact of remote learning.

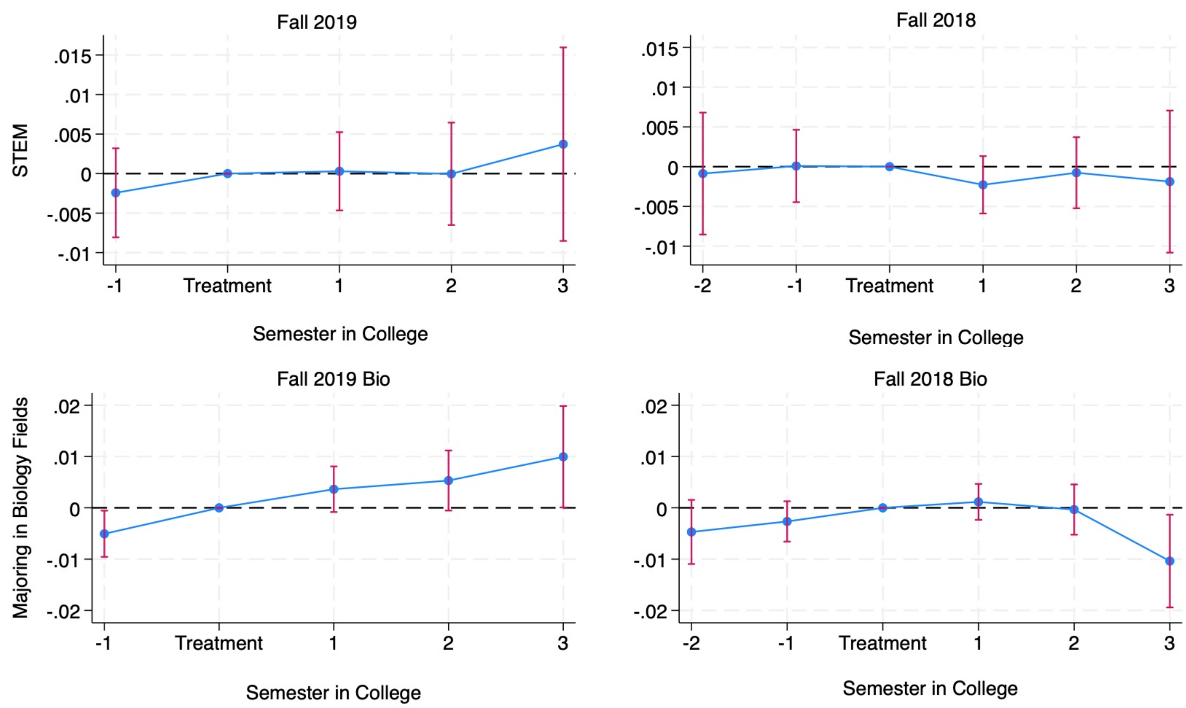


Figure 3.2: Event Study Estimates of Majoring in STEM and Biology Fields

Note: This figure displays event-study estimates of the probability of majoring in STEM and biology-related fields, centered around the semester in which remote learning due to COVID-19 began (Spring 2020). Each panel represents a separate cohort-by-major grouping. The vertical axis shows the difference between treated and control groups compared to the difference observed in the reference period (semester immediately before remote instruction began). The estimates are based on a Difference-in-Differences framework with staggered treatment timing, and show delta values from the DiD equation (Equation 3.2). The top row shows results for majoring in STEM (Fall 2019 only) and Biology (Fall 2018 cohort), while the bottom panels show probabilities of majoring in Biology for Fall 2018 and Fall 2019 students. Error bars represent 95% confidence intervals. Controls include gender, first-gen status and financial aid received.

The event-study estimates from the DiD specification are plotted in Figure 3.2. This figure displays estimated semester-by-semester treatment effects—along with 95% confidence intervals—on students' probability of majoring in STEM or Biology-related fields. The analysis uses a staggered treatment Difference-in-Differences framework, in which the timing of exposure to remote instruction (i.e., the COVID-19 shock) varies by student cohort.

For most outcomes, including majoring in STEM (top-left panel, Fall 2019) and Biology for the Fall 2018 cohort (bottom-right panel), the pre-treatment coefficients are close to zero and statistically insignificant, supporting the plausibility of the parallel trends assumption. This implies that, prior to the onset of remote learning, students across treatment and control groups were following similar academic trajectories in terms of major retention.

However, for the Fall 2019 cohort in Biology (bottom-left panel), I observe a notable upward linear trend in the pre-treatment period, suggesting that students in this group were already becoming increasingly likely to major in Biology prior to remote learning. This threatens the internal validity of causal inference for this subgroup, as it raises the possibility that post-treatment effects may be confounded by continuation of an existing trajectory rather than a discrete COVID-related shift.

Given there is suggestive evidence of pre-trends for Biology related fields, I re-estimate the event study under a linear trends assumption following recent methodological recommendations by Rambachan and Roth (2023) and Dette and Schumann (2024). I also pool the sample i.e. the sample includes students from Fall 2013 to Fall 2019. As shown in Figure 3.3, this adjustment removes any evidence of a post-treatment increase in biology enrollment. If anything, the direction of the effect reverses: the relative likelihood of majoring in biology declines slightly by the third post-pandemic semester.

To assess the impacts of the estimate, I compare the effect size based on mean values of STEM and Biology Related majors. As shown in Table 3.2, I compare the effect sizes by considering the mean values for STEM enrollment and biology related enrollment respectively. Even under the most extreme estimates within the confidence bounds (assume effect is 0.015 for STEM), the implied effect sizes remain under 0.035 of the mean for STEM (STEM mean=0.41). Assuming effect is -0.014 for Bio-related fields, the implied effect size is 0.07 of the mean (Bio-Related major mean=0.18). I therefore argue that—regardless of sign—the magnitude of the remote learning shock on biology enrollment is too small to be considered substantively meaningful. Although not shown, these results hold across alternative specifications (no controls, adding more controls, limiting sample of cohorts etc.) Thus, despite the scale and severity of the COVID-19 disruption, I find no evidence that remote instruction meaningfully altered the academic field-of-study trajectories of undergraduate students at the population level.

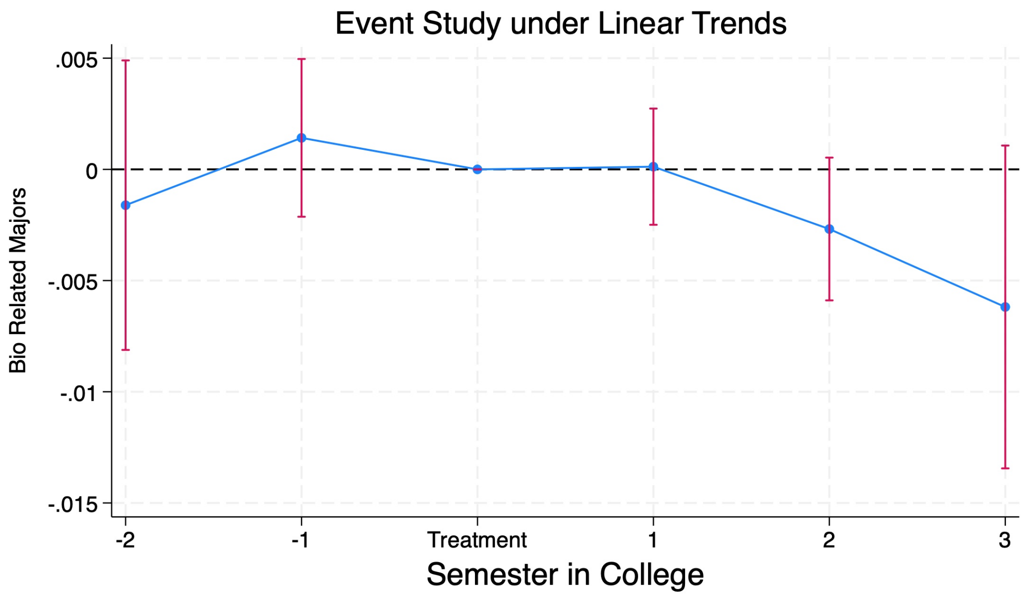


Figure 3.3: Event Study Estimates of Majoring in Biology Fields (Fall 2013–2019)

Note: This figure presents event-study estimates of the probability of majoring in biology-related fields, centered on the semester in which COVID-19-induced remote learning began (Spring 2020). The analysis pools all entry cohorts from Fall 2013 to Fall 2019 and implements a Difference-in-Differences framework with staggered treatment timing and assumes a linear trend model. The vertical axis shows the difference in outcomes between treated and not-yet-treated students relative to Spring 2020 (event time $t = 0$). Error bars denote 95% confidence intervals. Controls include gender, first-gen status and financial aid received.

The effect size is at most a 1 percentage point increase compared to a mean of 41% for enrollment as a STEM major. Assuming a linear trend, the effect size is at most a 1.5 percentage point decrease compared to a mean of 18% for enrollment in a biology-related field.

3.5.1 Effects on Switching into and out of STEM and Biology Majors

I next examine whether the COVID-19-induced transition to remote learning influenced students' likelihood of switching into or out of STEM and biology-related majors. Switching behavior serves as a useful indicator of academic realignment, especially during the early semesters of college when students frequently reassess their field-of-study decisions. If the pandemic disrupted advising systems, delayed lab-based experiences, or altered career priorities, one might expect changes in switching patterns.

As Table 3.2 shows, switching behavior is relatively uncommon. Only about 1.4% of students switch into a STEM major after starting in a non-STEM field, and fewer than 1% switch into biology-related fields. On the other hand, the mean share of student-semester observations classified as STEM majors is about 41.7%, and the mean for biology-related majors is around 18% (Table 3.2).

Figure 3.4 presents event-study estimates of the treatment effect on switching into STEM and biology majors, measured relative to the onset of remote learning in Spring 2020. The pandemic's arrival coincides with a small but detectable dip in switching behavior: the point estimate for switching into STEM majors falls by about 0.004 (0.4 percentage points), and for biology, by roughly 0.002. Although statistically significant at conventional levels, the effect does not last and the later event studies show that switching in behavior does not persist.

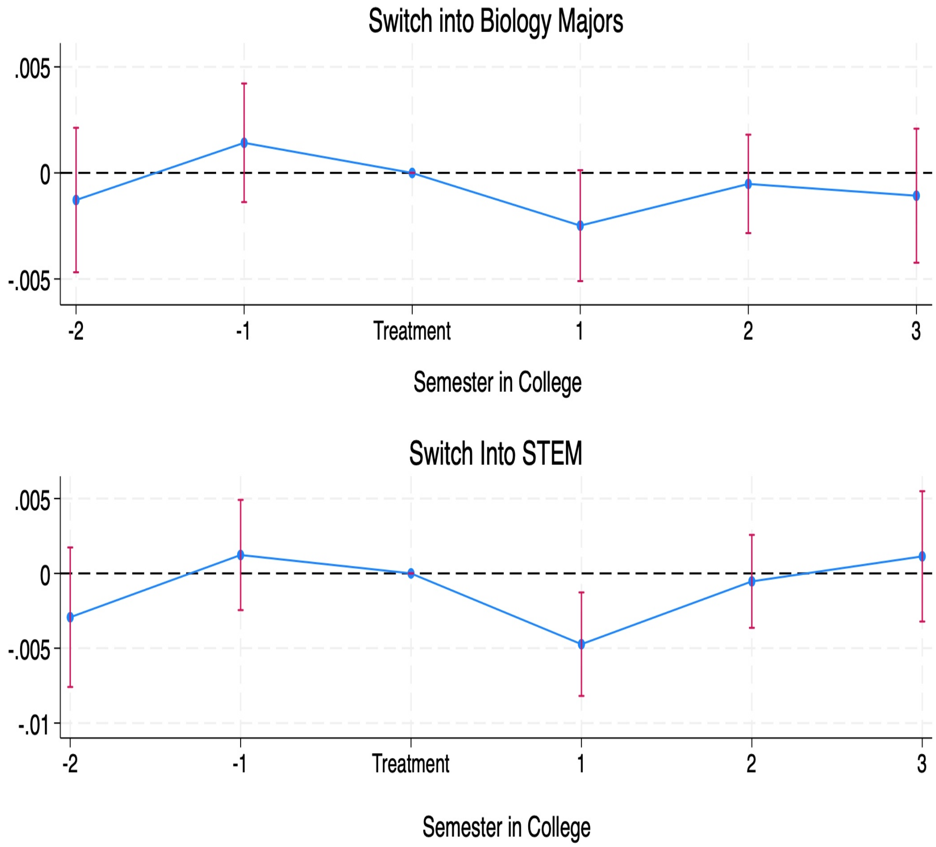


Figure 3.4: Event Study Estimates of Switching into STEM and Biology Fields

Note: This figure plots event-study estimates for switching into STEM (bottom panel) and biology-related (top panel) majors. The x-axis reflects time relative to the semester in which remote instruction began (Spring 2020). The estimates are from a Difference-in-Differences specification with group and time fixed effects, and the results reflect β_k coefficients from Equation 3.2. Error bars show 95% confidence intervals. Controls include gender, first-gen status, and financial aid.

Figure 3.5 presents the event-study estimates for students switching out of STEM and biology majors. The results indicate a small but noticeable decline in switching-out behavior during the first semester of remote instruction (Spring 2020). Prior to the treatment semester, eventually affected students exhibited slightly higher switching-out rates from bio-related majors, averaging approximately 0.005 units higher than their unaffected peers. However, this difference dissipates immediately upon the onset of remote learning, with estimates dropping to zero in the treatment semester. This decline suggests a potential anticipatory effect or immediate adjustment to the new learning environment. Importantly, the effect is transitory, as switching-out rates remain stable and centered around zero in subsequent semesters, indicating no sustained upward or downward trend in major-switching behavior post-treatment.

This pattern suggests that the initial disruption caused by the transition to remote instruction primarily delayed both entry into and exit from STEM and biology pathways, rather than generating lasting structural changes in students' academic trajectories. In other words, students were not more likely to switch majors at all—either into or out of STEM/Biology—during the pandemic semester, which may reflect temporary institutional delays or student hesitancy to make major academic decisions in a period of uncertainty.

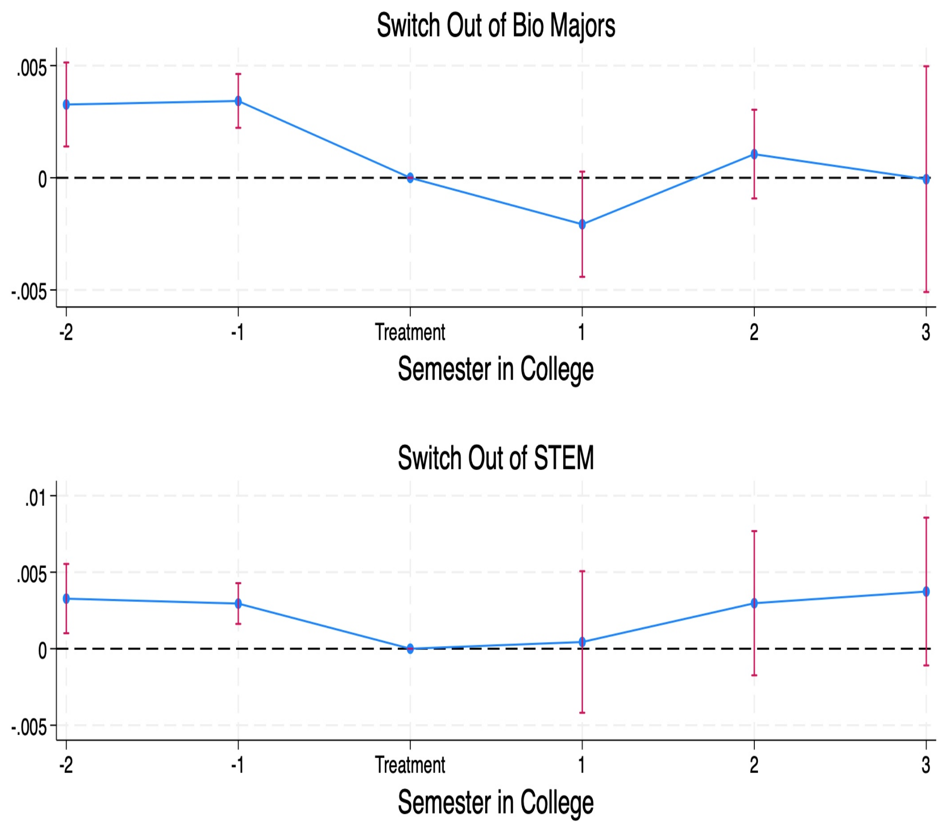


Figure 3.5: Event Study Estimates of Switching out of STEM and Biology Fields

Note: This figure plots event-study estimates for switching out of STEM (bottom panel) and biology-related (top panel) majors. The x-axis reflects time relative to the semester in which remote instruction began (Spring 2020). Estimates reflect β_k coefficients from Equation 3.2, with group and time fixed effects. Error bars are 95% confidence intervals. Controls include gender, first-gen status, and financial aid received.

One might argue that treatment effects appear more meaningful when framed relative to rare switching behaviors. For instance, a 0.005 increase in switching into STEM, given a base rate (mean) of 0.025, implies a 20% increase. However, this framing can be misleading. Because switching is infrequent to begin with, even sizable relative changes still correspond to small absolute movements in student behavior. Furthermore, base rates this low are often noisy, and percentage changes can be artificially inflated by small denominators. Thus, while there may be detectable shifts in switching behavior, the magnitude of change is too small to suggest any widespread re-sorting into or out of STEM or biology pathways.

In short, while some signals of academic reorientation exist, particularly in switching patterns, these are limited in scope. The results suggest that the pandemic and remote learning may have nudged the margins—but they did not materially alter the overall shape of the STEM or biology major pipelines in any meaningful way.

3.5.2 Disaggregation, Robustness Checks and Threats to Validity

To assess the credibility of the primary findings, I conduct a series of robustness checks to ensure that the estimated effects remain stable across alternative model specifications and identifying assumptions. First, although not shown here for brevity, I re-estimate the staggered Difference-in-Differences (DiD) models using alternative cohort definitions and a more comprehensive set of covariates. These include both time-varying and time-invariant controls such as semester GPA, financial aid amounts, high school academic characteristics, and key demographic indicators. Across all specifications, the estimated effects of remote instruction on STEM and biology major outcomes remain small in magnitude and often statistically insignificant. The consistency of these results across model variants reinforces the stability and credibility of the baseline estimates.

In a further set of analyses, I disaggregate the results by gender and first-generation college status. The effects remain directionally similar and statistically indistinguishable from zero in these subsamples, suggesting that the baseline findings are not driven by differential effects across these demographic groups. Due to limitations in statistical power, I do not further disaggregate the results by race or income group; estimates in those subsamples are considerably noisier. Future work with larger samples may be better positioned to explore these subgroup effects with greater precision.

I next turn to a complementary non-parametric robustness approach: nearest-neighbor matching. I implement nearest-neighbor matching following the bias-adjusted methodologies of Abadie et al. (2004) and Abadie et al. (2011). This method estimates the average treatment effect on the treated (ATET) by comparing COVID-affected students (i.e., those in the Fall 2018 and Fall 2019 entry cohorts) with pre-pandemic students who had similar observed characteristics but completed college before remote learning began. Students were matched using Mahalanobis distance across several covariates—high school GPA, standardized tests, financial aid, first-generation status, and semester of enrollment—to ensure comparability on pre-treatment characteristics. The matched design provides an alternative identification strategy that relaxes parametric modeling assumptions while allowing for intuitive estimation of counterfactual outcomes.

Table 3.5 summarizes the ATET estimates for four outcomes: (1) ultimately majoring in STEM, (2) majoring in a biology-related field, (3) switching into STEM, and (4) switching into biology-related majors. Given that behaviors in biology appeared to vary markedly across different cohorts, for likelihood of majoring in bio-related fields I use the full sample, whereas for STEM outcomes I restrict analysis to students entering in Fall 2018 or Fall 2019 compared to Fall 2016 and Fall 2017 cohorts.

Table 3.5: Nearest-Neighbor Matching Estimates of Remote Learning

Outcome	**ATET**	**Std. Error**	**95% CI**	N	**Min**	**Max**
STEM	0.0019	(0.009)	$[-0.017, 0.021]$	19,247	10	260
Switch STEM	-0.0065^{**}	(0.029)	$[-0.012, -0.007]$	19,335	10	265
Biology	0.0142^{***}	(0.004)	$[-0.007, 0.022]$	69,800	10	80
Switch Biology	-0.0019^{***}	(0.0006)	$[-0.003, -0.0007]$	69,800	10	80

Note: Each row reports the average treatment effect on the treated (ATET) from a nearest-neighbor matching model for the specified outcome. The treated group consists of students who experienced the COVID-19 remote instruction period during their undergraduate studies (e.g., Fall 2018 and Fall 2019 entry cohorts). Control students are drawn from earlier cohorts who did not face the disruption. Matching is based on Mahalanobis distance using pre-pandemic covariates, including high school GPA, financial aid, and first-generation status. Each student is exactly matched on the semester of observation. "Min Matches" and "Max Matches" refer to the minimum and maximum number of matches found per treated observation. Given that behaviors in biology appeared to vary markedly across different cohorts, for likelihood of majoring in bio-related fields I use the full sample, whereas for STEM outcomes I restrict analysis to students entering in Fall 2018 or Fall 2019 compared to Fall 2016 and Fall 2017 cohorts.

Consistent with the DiD estimates, the matching results show no statistically significant change in STEM majoring overall (ATET = 0.0019). I do, however, observe negative and statistically significant effects (at the 95% confidence level) for switching into STEM and for switching into biology-related majors, albeit of relatively small magnitudes. When comparing the matching estimates to the corresponding standard deviations (from Table 3.2), even the upper and lower bound estimates rarely exceed 0.03–0.05 standard deviations.

Threats to Validity

While the event-study estimates do not reveal any sustained or substantial shifts in STEM or biology major persistence following the transition to remote instruction, one potential threat to validity is selective attrition. Specifically, students who dropped out of college during or after the pandemic may systematically differ from those who remained enrolled. If students more likely to major in STEM disproportionately dropped out after the onset of COVID-19, the main estimates could be biased downward, as the continued participation or switching behavior of these students would not be observed.

This concern is particularly relevant because the sample is restricted to students who remained enrolled through at least the end of their fourth academic year. If remote learning disproportionately caused certain types of students—especially those enrolled in STEM majors—to drop out, the estimated effects on major switching or persistence could be biased downward, potentially attenuating true treatment effects. To explicitly investigate whether the COVID period had a differential impact on dropout risk for STEM and biology majors, I analyze an expanded sample that includes all students who dropped out at any point within their first six semesters (excluding those who graduated early). Using this broader sample, I estimate a student-level fixed effects regression model, where the dependent variable is a binary indicator capturing whether a student drops out in a given term. The primary explanatory variables include lagged indicators for enrollment in STEM or biology majors, with additional controls for lagged GPA, semester tenure indicators, and a dummy variable denoting the pandemic period.

Due to potential collinearity between STEM and biology majors, these indicators are included in separate models. Including both STEM and biology in the same specification would obscure the independent effects of each major and potentially lead to misleading inferences. By estimating separate models, I ensure that the interpretation of each coefficient remains distinct and unbiased. The specification is as follows:

$$\begin{aligned}\text{Dropout}_{it} = \alpha_i &+ \beta_1 \text{STEM/Bio}_{i,t-1} + \beta_2 \text{COVID}_t \\ &+ \beta_3(\text{STEM/Bio}_{i,t-1} \times \text{COVID}_t) + \beta_4 \text{GPA}_{i,t-1} \\ &+ \sum_{s=2}^{t} \gamma_s \mathbf{1}[\text{Semester}_t = s] + \varepsilon_{it}\end{aligned} \tag{3.5}$$

The interaction term β_3 allows for a more nuanced understanding of whether the COVID period had a differential effect on dropout probabilities for students in STEM or biology majors. A corresponding probit model is specified as:

$$\begin{aligned}\Pr(\text{Dropout}_{it} = 1 \mid X_{it}, u_i) = \Phi\Big(&\beta_1 \text{STEM/Bio}_{i,t-1} + \beta_2 \text{COVID}_t \\ &+ \beta_3(\text{STEM/Bio}_{i,t-1} \times \text{COVID}_t) \\ &+ \beta_4 \text{GPA}_{i,t-1} \\ &+ \sum_{s=2}^{t} \gamma_s \mathbf{1}[\text{Semester}_t = s] + u_i\Big)\end{aligned} \tag{3.6}$$

The probit model incorporates random effects to account for unobserved heterogeneity across students. Unlike fixed effects, which cannot be directly estimated in non-linear models due to the incidental parameters problem, random effects are more suitable in the probit framework (StataCorp, 2025).

Overall, the findings provide reassurance regarding the potential attrition bias concern raised earlier. Specifically, the interaction terms between major enrollment and the COVID period are statistically insignificant across all model specifications (Table 3.6). This result suggests that the pandemic did not disproportionately increase dropout rates among STEM or biology majors relative to other students. Additionally, the negative and statistically significant coefficient for lagged STEM enrollment implies that STEM majors were slightly less likely to drop out overall, further reducing the concern that selective attrition could have driven the null findings on major switching or persistence. Therefore, the observed patterns in switching behavior are unlikely to be substantially biased downward due to differential attrition across STEM or biology majors during the COVID period.

Table 3.6: Lagged Major Enrollment and Dropout Risk: Interaction Effects with COVID Period

	FE (STEM)	**FE (Biology)**	**Probit (STEM)**	**Probit (Biology)**
Lagged STEM Major	-0.0506^{***}	–	-0.0759^{***}	–
	(0.0130)		(0.0175)	
Lagged Biology Major	–	-0.0130	–	-0.0286
		(0.0173)		(0.0214)
COVID Semester	-0.0011	-0.0052	-0.0238	-0.0338
	(0.0132)	(0.0120)	(0.0263)	(0.0233)
Interaction (STEM × COVID)	-0.0170	–	-0.0150	–
	(0.0192)		(0.0410)	
Interaction (Bio × COVID)	–	-0.0096	–	0.0152
		(0.0247)		(0.0523)
Lagged Term GPA	-0.1549^{***}	-0.1537^{***}	-0.1909^{***}	-0.1897^{***}
	(0.0059)	(0.0059)	(0.0063)	(0.0063)
Semester Fixed Effects	Yes	Yes	Yes	Yes
Student Effects	Fixed	Fixed	Random	Random
Observations	30,912	30,912	30,912	30,912
Individuals	10,046	10,046	10,046	10,046

Note: Table reports estimates from fixed-effects linear regressions and random-effects probit models predicting dropout in a given semester. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Interaction terms assess whether the COVID period had a differential impact on dropout risk for STEM and biology majors. Random effects in probit models account for unobserved heterogeneity while allowing for time-invariant covariates (StataCorp, 2025).

3.6 Conclusion

This study examined whether the COVID-19-induced transition to remote instruction meaningfully altered students' decisions to remain in or switch into STEM and biology-related majors. Leveraging longitudinal administrative data from a large U.S. public university and employing a staggered Difference-in-Differences design, I find that remote learning had negligible effects on students' academic field-of-study choices. Across all specifications—including event-study models, fixed effects regressions, and nearest-neighbor matching—the estimated impacts on STEM or biology majoring and switching behavior were small, statistically insignificant in most cases, and economically trivial.

While majoring in biology and health related majors showed a modest apparent increase during remote semesters for the Fall 2019 cohort, this pattern disappeared once I accounted for underlying cohort-specific linear trends. In fact, after adjustment, the point estimates for biology enrollment turn slightly negative. For switching behavior, I observe a brief dip in transitions into STEM and biology fields during the immediate aftermath of the shift to remote learning—particularly in Fall 2020—but these disruptions dissipate within one to two semesters. The longer-run pattern shows full recovery, with no evidence of permanent behavioral change.

These findings suggest that the abrupt transition to remote instruction—despite being an unprecedented shock to higher education—did not substantially reshape students' academic trajectories, at least in terms of their major field of study. Students who remained enrolled during the pandemic demonstrated striking stability in their educational paths. Contrary to anecdotal accounts predicting large-scale attrition from lab-based majors or a boom in interest in biomedical disciplines, the evidence points to broad continuity. The mode of instruction, at least in the short run, did not fundamentally alter students' academic intentions in STEM or biology-related fields.

Importantly, while I control for key student characteristics—such as gender, first-generation status, and financial need—through both regression adjustment and exact matching procedures, I do not disaggregate treatment effects by race, or income group. The institutional dataset, while large overall, produces very limited sample sizes for some subgroups when stratified by cohort and semester, leading to prohibitively noisy estimates in subgroup analyses. Although preliminary checks suggest that treatment effects are directionally consistent across these subpopulations, I cannot make strong statistical claims about heterogeneity due to lack of statistical power. Future research using larger multi-institutional datasets or national administrative records would be better positioned to assess differential impacts of remote instruction on specific student groups.

Several other limitations should also be noted. First, the data are drawn from a single large public university, and the external validity of the findings is therefore limited. Institutions with different instructional capacities, advising infrastructure, or student demographics might experience different responses to remote learning. Second, the analysis captures only short-run effects—namely, majoring and switching behavior in the semesters immediately following the onset of the pandemic. I do not observe longer-term outcomes such as STEM degree completion, graduate school attendance, or labor market entry. Finally, although the staggered DiD strategy incorporates cohort and time fixed effects and controls for key confounders, I cannot rule out the possibility of residual bias from unobserved, time-varying shocks that coincided with the pandemic—such as changes in family income, health conditions, or macroeconomic expectations—that may independently influence major decisions.

In sum, the evidence from this study indicates that the transition to remote learning during COVID-19 had, at most, marginal and short-lived effects on undergraduate students' decisions to persist in or switch into STEM and biology-related majors. This empirical stability highlights the resilience of students' academic preferences even amid substantial disruption to instructional delivery. While the pandemic undoubtedly transformed the delivery of higher education, the findings suggest that it did not fundamentally alter students' academic ambitions. These results offer a data-driven counterpoint to early speculative narratives forecasting either mass attrition from STEM or a widespread pivot toward biomedical fields. Instead, they provide a baseline for understanding the limited scope of behavioral response to even large-scale instructional shocks. That said, the study does not address whether students learned less, acquired fewer skills, or experienced diminished returns to STEM coursework during the remote instruction period. Understanding whether the format of instruction affected the quality of learning—rather than just enrollment patterns—remains a critical open question for future research.

Long story short, while the delivery of education changed overnight, the direction of student ambition did not. That distinction matters—and it underscores a deeper insight: structural shocks may shake the system, but they do not easily rewrite student intent.

Bibliography

Abadie, A., Drukker, D., Herr, J. L., & Imbens, G. W. (2004). Implementing matching estimators for average treatment effects in stata. *The Stata Journal*, *4*(3), 290–311.

Abadie, A., Drukker, D., Herr, J. L., & Imbens, G. W. (2011). Bias-corrected matching estimators for average treatment effects. *Journal of Business & Economic Statistics*, *29*(1), 1–11. https://doi.org/10.1198/jbes.2009.07333

Angrist, J. D., & Lavy, V. (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, *114*(2), 533–575.

Angrist, J. D., Lavy, V., Leder-Luis, J., & Shany, A. (2019). Maimonides rule redux. *American Economic Review: Insights*, *1*(3), 309–324.

Arcidiacono, P., Aucejo, E. M., & Hotz, V. J. (2016). University differences in the graduation of minorities in stem fields: Evidence from california. *American Economic Review*, *106*(3), 525–562. https://doi.org/10.1257/aer.20130626

Ardington, C., Wills, G., & Kotze, J. (2021). Covid-19 learning losses: Early grade reading in south africa. *International Journal of Educational Development*, *86*, 102480.

Aucejo, E. M., French, J., Araya, M. P. U., & Zafar, B. (2020). The impact of covid-19 on student experiences and expectations: Evidence from a survey. *Journal of Public Economics*, *191*, 104271. https://doi.org/10.1016/j.jpubeco.2020.104271

Azevedo, J. P., Hasan, A., Goldemberg, D., Geven, K., & Iqbal, S. A. (2021). Simulating the potential impacts of covid-19 school closures on schooling and learning outcomes: A set of global estimates. *The World Bank Research Observer*, *36*(1), 1–40.

Bacher-Hicks, A., Goodman, J., & Mulhern, C. (2021). Inequality in household adaptation to schooling shocks: Covid-induced online learning engagement in real time. *Journal of Public Economics*, *193*, 104345.

Bailey, M. J., & Dynarski, S. M. (2011). Inequality in postsecondary attainment. *Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low-Income Children*, *117*, 117–132.

Barber, P. H., Shapiro, C., Jacobs, M. S., Avilez, L., Brenner, K. I., Cabral, C., Cebreros, M., Cosentino, E., Cross, C., Gonzalez, M. L., Lumada, K. T., Menjivar, A. T., Narvaez, J., Olmeda, B., Phelan, R., Purdy, D., Salam, S., Serrano, L., Velasco, M. J., ... Levis-Fitzgerald, M. (2021). Disparities in remote learning faced by first-generation and underrepresented minority students during covid-19: Insights and opportunities from a remote research experience. *Journal of Microbiology & Biology Education*, *22*(1), 10.1128/jmbe.v22i1.2457. https://doi.org/10.1128/jmbe.v22i1.2457

Barnett, E. A., Bork, R. H., Mayer, A. K., Pretlow, J., Wathington, H. D., & Weiss, M. J. (2012). Bridging the gap: An impact study of eight developmental summer bridge programs in texas. *National Center for Postsecondary Research*.

Baron, E. J. (2018). The effect of teachers' unions on student achievement in the short run: Evidence from wisconsin's act 10. *Economics of Education Review*, *67*, 40–57.

Belley, P., & Lochner, L. (2007). Changes in post-secondary education attainment across cohorts: The role of parental education. *Journal of Labor Economics*, *25*(3), 463–498.

Bettinger, E. P., & Long, B. T. (2009). Addressing the needs of underprepared students in higher education: Does college remediation work? *Journal of Human Resources*, *44*(3), 736–771. https://doi.org/10.3368/jhr.44.3.736

Biasi, B. (2021). The labor market for teachers under different pay schemes. *American Economic Journal: Economic Policy*, *13*(3), 63–102. https://doi.org/10.1257/pol.20200295

Biasi, B., Fu, C., & Stromme, J. (2021). *Equilibrium in the market for public school teachers: District wage strategies and teacher comparative advantage* (tech. rep. No. w28530). National Bureau of Economic Research.

Bird, K. A., Castleman, B. L., & Lohner, G. (2022). Negative impacts from the shift to online learning during the covid-19 crisis: Evidence from a statewide community college system. *AERA Open*, *8*. https://doi.org/10.1177/23328584221081220

Branson, Z., & Mealli, F. (2018). The local randomization framework for regression discontinuity designs: A review and recommendations for practitioners. *arXiv preprint arXiv:1810.02761*.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Cabrera, N. L., Miner, D. D., & Milem, J. F. (2013). Can a summer bridge program impact first-year persistence and performance?: A case study of the new start summer program. *Research in Higher Education*, *54*, 481–498.

Cahalan, M., Perna, L. W., Yamashita, M., Wright-Kim, J., & Jiang, N. (2021). Trends in pell grant receipt and postsecondary outcomes. *Pell Institute for the Study of Opportunity in Higher Education*.

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Callaway, B., Goodman-Bacon, A., & Sant'Anna, P. H. (2024). *Difference-in-differences with a continuous treatment* (tech. rep.). National Bureau of Economic Research.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge University Press.

Card, D., Lee, D. S., Pei, Z., & Weber, A. (2017). Regression kink design: Theory and practice. In *Regression discontinuity designs: Theory and applications* (pp. 341–382). Emerald Publishing Limited.

Castleman, B. L., & Page, L. C. (2015). Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates? *Journal of Economic Behavior & Organization*, *115*, 144–160.

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2024). *Practical introduction to regression discontinuity designs: Volume ii*. Cambridge University Press.

Cattaneo, M. D., Titiunik, R., & Vazquez-Bare, G. (2016). Inference in regression discontinuity designs with a discrete running variable. *Stata Journal*, *16*(1), 331–367.

Chen, X., Zou, D., Xie, H., & Wang, F. L. (2021). Past, present, and future of smart learning: A topic-based bibliometric analysis. *International Journal of Educational Technology in Higher Education*, *18*, 2–29. https://doi.org/10.1186/s41239-020-00239-6

Cunial, S. L. (2021). Transitions for whom? political alignment and subsidies for solar energy projects in rural colombian municipalities. *Latin American Policy*, *12*, 300–332. https://doi.org/10.1111/lamp.12219

De Chaisemartin, C., & d'Haultfoeuille, X. (2024). Difference-in-differences estimators of intertemporal treatment effects. *Review of Economics and Statistics*, 1–45.

de Chaisemartin, C., & D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, *110*(9), 2964–2996. https://doi.org/10.1257/aer.20181169

Dette, H., & Schumann, M. (2024). Testing for equivalence of pre-trends in difference-in-differences estimation. *Journal of Business & Economic Statistics*, *42*(4), 1289–1301. https://doi.org/10.1080/07350015.2024.2308121

Dube, A., Girardi, D., Jordà, Ò., & Taylor, A. M. (2023). *A local projections approach to difference-in-differences* (NBER Working Paper No. 31184). National Bureau of Economic Research. https://doi.org/10.3386/w31184

Dynarski, S. (2003). Does aid matter? measuring the effect of student aid on college attendance and completion. *American Economic Review*, *93*(1), 279–288.

Dynarski, S., Nurshatayeva, A., Page, L. C., & Scott-Clayton, J. (2023). Addressing nonfinancial barriers to college access and success: Evidence and policy implications. In *Handbook of the economics of education* (pp. 319–403, Vol. 6). Elsevier.

Eberts, R. W., & Stone, J. A. (1988). Student achievement in public schools: Do principals make a difference? *Economics of Education Review*, *7*(3), 291–299. https://doi.org/10.1016/0272-7757(88)90019-3

Ersoy, F., & Speer, J. D. (2025). Opening the black box of college major choice: Evidence from an information intervention. *Journal of Economic Behavior & Organization*, *231*, 106800. https://doi.org/10.1016/j.jebo.2024.106800

Ersoy, F. Y. (2020). The effects of the great recession on college majors. *Economics of Education Review*, *77*, 102018. https://doi.org/10.1016/j.econedurev.2020.102018

Fryer, R. G. (2013). Teacher incentives and student achievement: Evidence from new york city public schools. *Journal of Labor Economics*, *31*(2), 373–407. https://doi.org/10.1086/667757

Goodman, S. F., & Turner, L. J. (2013). The design of teacher incentive pay and educational outcomes: Evidence from the new york city bonus program. *Journal of Labor Economics*, *31*(2), 409–420. https://doi.org/10.1086/667846

Goodman-Bacon, A. (2021a). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277. https://doi.org/10.1016/j.jeconom.2021.03.014

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

Griffith, A. L. (2010). Persistence of women and minorities in stem field majors: Is it the school that matters? *Economics of Education Review*, *29*(6), 911–922. https://doi.org/10.1016/j.econedurev.2010.06.010

Han, E., & Garcia, E. (2022). Teachers' base salary and districts' academic performance: Evidence from national data. *Sage Open*. https://doi.org/10.1177/21582440221082138

Hanushek, E. A., Piopiunik, M., & Wiederhold, S. (2019). The value of smarter teachers: International evidence on teacher cognitive skills and student performance. *Journal of Human Resources*, *54*(4), 857–899. https://doi.org/10.3368/jhr.54.4.0617.8961R1

Henry, G. T., & Redding, C. (2020). The consequences of leaving school early: The effects of within-year and end-of-year teacher turnover. *Education Finance and Policy*, *15*(2), 332–356. https://doi.org/10.1162/edfp_a_00274

Institutional Research, Planning, and Effectiveness. (2021). Qualitative study on why students leave csu [Accessed: 2024-08-11]. https://www.ir.colostate.edu/qualitative-study-on-why-students-leave-csu/

Jack, R., Halloran, C., Okun, J., & Oster, E. (2023). Pandemic schooling mode and student test scores: Evidence from us school districts. *American Economic Review: Insights*, *5*(2), 173–190.

Jackson, C. K., Johnson, R. C., & Persico, C. (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *The Quarterly Journal of Economics*, *131*(1), 157–218. https://doi.org/10.1093/qje/qjv036

Jackson, C. K., & Mackevicius, C. L. (2024). What impacts can we expect from school spending policy? evidence from evaluations in the united states. *American Economic Journal: Applied Economics*, *16*(1), 412–446. https://doi.org/10.1257/app.20220279

Kitchen, J. A., Sonnert, G., & Sadler, P. M. (2018). Effect of a stem talent expansion program on undergraduates' career aspirations. *Journal of STEM Education: Innovations and Research*, *19*(3), 24–30.

Koopmann, J., Zimmer, L. M., & Lörz, M. (2024). The impact of covid-19 on social inequalities in german higher education. an analysis of dropout intentions of vulnerable student groups. *European Journal of Higher Education*, *14*(2), 290–307.

Ladd, H. F., & Sorensen, L. C. (2017). Returns to teacher experience: Student achievement and motivation in middle school. *Education Finance and Policy*, *12*(2), 241–279. https://doi.org/10.1162/EDFP_a_00194

Law, T. (2021). I truly did find my calling. meet the young people shaping health care's post-pandemic future [Time, May 27. Retrieved from https://time.com/6053600/pandemic-health-care-students/].

Leuven, E., & Løkken, S. A. (2020). Long-term impacts of class size in compulsory school. *Journal of Human Resources*, *55*(1), 309–348.

Levanon, G. (2023). How changing college majors is reshaping the future [LinkedIn, Retrieved from https://www.linkedin.com/pulse/how-changing-college-majors-reshaping-future-gad-levanon-hhooc/].

Li, F., Mattei, A., & Mealli, F. (2015). Evaluating the causal effect of university grants on student dropout: Evidence from a regression discontinuity design using principal stratification. *The Annals of Applied Statistics*, *9*(4), 1906–1931.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomforest. *R news*, *2*(3), 18–22.

Litschwartz, S. (2022). Local randomization regression discontinuity designs when test scores are the running variable. *arXiv preprint arXiv:2209.11558*. https://arxiv.org/abs/2209.11558

Long, B. T. (2014). The financial crisis and college enrollment: How have students and their families responded? (J. R. Brown & C. M. Hoxby, Eds.), 209–233. https://doi.org/10.7208/chicago/9780226201979.003.0007

Miller, A. (2013). Principal turnover and student achievement. *Economics of Education Review*, *36*, 60–72. https://doi.org/10.1016/j.econedurev.2013.05.004

Mostafa, S. A., Ferguson, R., Tang, G., & Ashqer, M. (2023). An analysis of the covid-19-induced flexible grading policy at a public university. *Higher Education Policy*, 1.

Muralidharan, K., Pradhan, M., & Rogers, H. (2016). Double for nothing? https://www.nber.org/papers/w21806

Muralidharan, K., & Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from india. *Journal of Political Economy*, *119*(1), 39–77.

Murphy, T. E., Gaughan, M., Hume, R., & Moore Jr, S. G. (2010). College graduation rates for minority students in a selective technical university: Will participation in a summer bridge program contribute to success? *Educational evaluation and policy analysis*, *32*(1), 70–83.

National Center for Education Statistics (NCES). (2017). Percentage of 2011–12 first time postsecondary students who had ever declared a major in an associate's or bachelor's degree program within 3 years of enrollment, by type of degree program and control of first institution: 2014 [Accessed March 2025]. https://nces.ed.gov/Datalab/TablesLibrary/TableDetails/11764

New Jersey Department of Education. (2024). School performance reports [Accessed: 2024-08-23]. https://www.nj.gov/education/doedata/fact.shtml

New Jersey Department of the Treasury. (2011). Pension reform 2011 [[Accessed: 2024-08-05]].

Ng, K. (2024). The effects of teacher tenure on productivity and selection. *Economics of Education Review*, *101*, 102558.

Olmedo-Cifuentes, I., & Martínez-León, I. M. (2022). University dropout intention: Analysis during covid-19. *Journal of Management and Business Education*, *5*(2), 97–117.

Oreopoulos, P., Brown, R. S., & Lavecchia, A. M. (2017). Pathways to education: An integrated approach to helping at-risk high school students. *Journal of political economy*, *125*(4), 947–984.

Page, L. C., & Scott-Clayton, J. (2016). Framing the decision to go to college: The role of information and context in college decision-making. *Economics of Education Review*, *51*, 68–73.

Page, P. (2023). The pandemic generation is taking its supply chain lessons to school [Wall Street Journal, Retrieved from https://www.wsj.com/articles/the-pandemic-generation-is-taking-its-supply-chain-lessons-to-school-45f0f8df].

Pertold-Gebicka, B. (2024). Medium-run effects of covid-19 induced distant learning on students' academic performance. *Labour Economics*, *89*, 102601.

Prudencio, D., Balmori-de-la-Miyar, J., Silverio-Murillo, A., & Sobrino, F. (2024). Examining covid-19's disruptive effect on education in mexican universities. *International Journal of Educational Development*, *111*, 103144. https://doi.org/10.1016/j.ijedudev.2024.103144

Rambachan, A., & Roth, J. (2020). An honest approach to parallel trends. *Working Paper*. https://github.com/asheshrambachan/HonestDiD

Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. *The Review of Economic Studies*, *90*(5), 2555–2591. https://doi.org/10.1093/restud/rdad018

Rodriguez-Planas, N. (2022). Hitting where it hurts most: Covid-19 and low-income urban college students. *Economics of Education Review*, *87*, 102233.

Rodríguez-Planas, N. (2022). Covid-19, college academic performance, and the flexible grading policy: A longitudinal analysis. *Journal of Public Economics*, *207*, 104606.

Roksa, J., & Kinsley, P. (2019). The role of family support in facilitating academic success of low-income students. *Research in Higher Education*, *60*, 415–436.

Shakya. (2025). *Evaluating small-scale scholarships at csu: Preliminary evidence from rdd and did designs* [Unpublished working paper, Colorado State University].

Springer, M. G., Ballou, D., Hamilton, L., Le, V.-N., Lockwood, J. R., McCaffrey, D. F., Pepper, M., & Stecher, B. M. (2012). Teacher pay for performance: Experimental evidence from the project on incentives in teaching (point). *Economics of Education Review*, *31*(1), 1–14. https://doi.org/10.1016/j.econedurev.2011.10.003

StataCorp. (2025). *Stata 19 longitudinal-data/panel-data reference manual* [Release 19]. College Station, TX, Stata Press. https://www.stata.com/bookstore/longitudinal-panel-data-reference-manual/

Stewart, S., Lim, D. H., & Kim, J. (2015). Factors influencing college persistence for first-time students. *Journal of Developmental Education*, 12–20.

Swartzentruber, D., & Kaizar, E. (2022). Small study regression discontinuity designs: Density inclusive study size metric and performance. *arXiv preprint arXiv:2209.01396*. https://arxiv.org/abs/2209.01396

What Works Clearinghouse (WWC). (2024). Summer bridge programs [https://ies.ed.gov/ncee/wwc/Intervention/824].

Xie, T. (2021). Work and covid-19: How the pandemic changed students' career plans [Teen Vogue, March 15. Retrieved from https://www.teenvogue.com/story/career-changes-covid-pandemic].

Ye, X., Zhai, M., Feng, L., Xie, A., Wang, W., & Wu, H. (2022). Still want to be a doctor? medical student dropout in the era of covid-19. *Journal of Economic Behavior & Organization*, *195*, 122–139.

Appendix Chapter 1

Compensation Glossary

- **Average Salary** – the base salary cost divided by the total number of full-time employees (FTE) on the scattergram.
- **Base Salary Cost** – The total of each step on the guide multiplied by each corresponding step on the scattergram. Other amounts may or may not be included, such as longevity, ratio differentials, extra-curricular activities, stipends, black seal amounts, building stipends, etc.
- **Breakage** – The amount of dollars saved between the salary of a departing employee (retirement, resignation, and leave of absence) and the new employee who is replacing the departed employee.

  Example:

  - $50,000 salary of retiring employee
  - $30,000 replacement employee
  - $20,000 breakage
- **Bubble/Balloon** – An abnormal separation between two steps on a salary guide.

  Example:

  - Step 13 - $39,000
  - Step 14 - $40,000
  - Maximum - $50,000
  - Increment - $10,000 or 25%
- **Cumulative Earnings** – The total sum of all salaries in a specified time period or career. NJEA Research calculates the 10-, 20-, and 30-year earnings based on a long-standing formula of 5 years on the BA column and the remaining years on the MA

column. Longevity is added, as are any other negotiated amounts at the appropriate time.

- **Guide Movement** – A movement from one step on a guide to the next higher step on that guide. Horizontal movement would be movement to a higher credit/degree/level guide based on a specified criteria.

- **Horizontal/Lane/Column** – A specific list of salaries with a minimum, maximum, and number of steps.

- **Increment** – The dollar amount of the salary increase an individual receives when they advance a step on the guide.

  Example:

  - Step 1 - $50,000
  - Step 2 - $51,000
  - Increment - $1,000

- **Increment Cost** – The dollar amount of the increment multiplied by the number of individuals that will receive that increment for a contract year.

- **Longevity** – Additional money paid to an employee above the salary guide. It is usually based on years of service to either the school district or the profession in general. It is usually a specified dollar amount, but can also be a percent of salary.

  Example:

  - $1000 additional for 15 years of service to the district
  - or 3% of individual salary for 15 years of service to the district.

- **Maximum** – The highest step on the published salary guide. It may also be called the career rate.

- **Minimum** – The beginning step of a guide that is considered to be the hiring step with no experience.
- **Off Guide Salaries** – Additional salaries that are paid above the printed salary guides. They are actually additional steps on a guide.
- **Salary Guide** – A chart that shows the dollar value of each step and level/category.
- **Scattergram** – A chart showing the number of employees on each step and level/category of a salary guide. These employees will generally be in the full-time equivalency (FTE) category of employment.

Table A1: Year Negotiation was Approved

Year	Frequency	Percent	Cumulative Percent
Never Treated	39	2.31%	2.31%
2006	9	0.53%	2.85%
2007	14	0.83%	3.68%
2008	64	3.80%	7.47%
2009	144	8.54%	16.01%
2010	204	12.10%	28.11%
2011	248	14.71%	42.82%
2012	157	9.31%	52.14%
2013	149	8.84%	60.97%
2014	92	5.46%	66.43%
2015	101	5.99%	72.42%
2016	86	5.10%	77.52%
2017	105	6.23%	83.75%
2018	80	4.74%	88.49%
2019	85	5.04%	93.53%
2020	44	2.61%	96.14%
2021	39	2.31%	98.46%
2022	4	0.24%	98.70%
2023	21	1.25%	99.94%
2024	1	0.06%	100.00%

Note: Since the school-level outcomes are from 2003–2019, I have a decent number of schools acting as controls. "Never treated" indicates that 39 schools have yet to adopt this policy. Year 2006 denotes that the negotiation was approved to go into effect in the 2006–2007 school year.

Table A2: Summary Statistics for Educational Outcomes

Variable	Observations	Mean	Std. Dev.	Min	Max
Math 4	17,945	70.77	21.69	5.6	100
Math 8	8,187	59.53	24.44	3.5	100
Math HS	5,203	67.59	26.01	7.4	100
ELA 4	13,848	70.96	18.79	10	100
ELA 8	9,587	72.98	20.17	10	100
ELA HS	5,134	75.79	22.72	10	100
HS Graduation	5,189	89.48	12.12	4.6	100
College Graduation	4,589	81.18	12.02	16.6	100

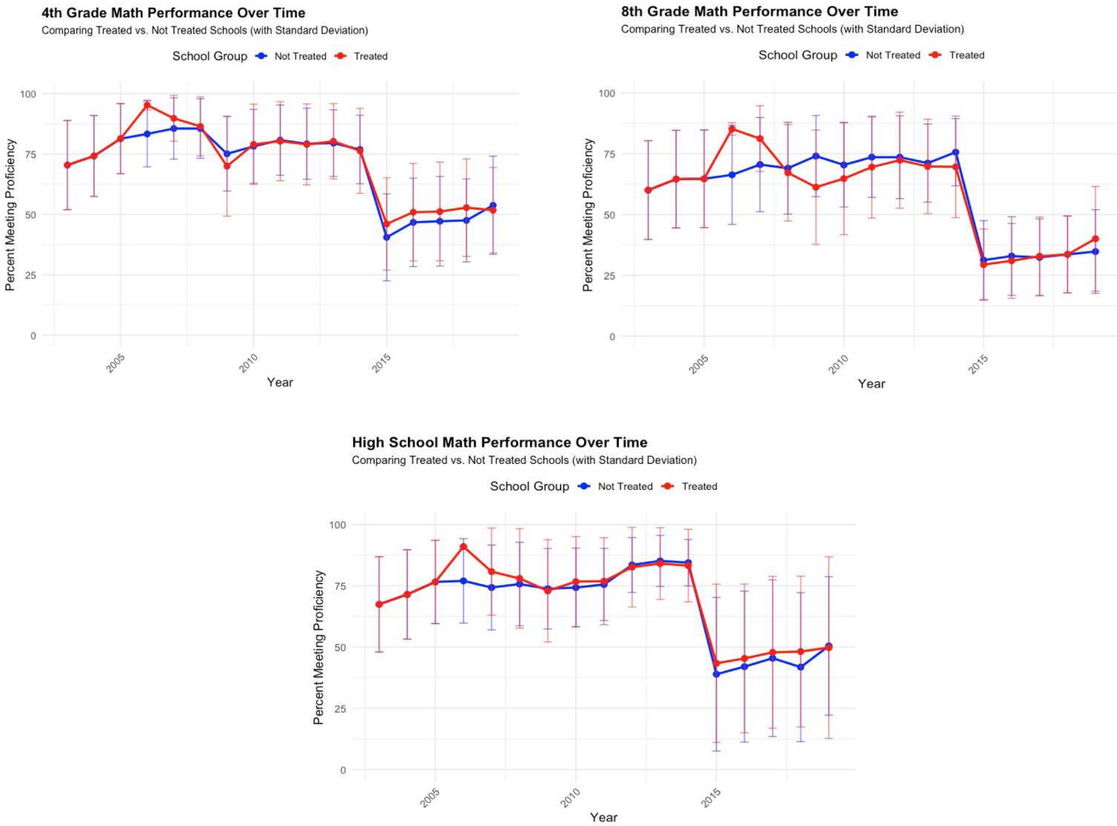


Figure A1: Evolution of Math Scores by Grade level

Note: This figure presents the average math scores for treated and not treated schools over time. "Not Treated" includes both schools that have never implemented the minimum salary policy and those that had not yet done so in a given year. "Treated" schools are those that had already implemented the policy. Because treatment adoption is staggered over time, simple mean comparisons do not denote causal effects. Standard deviation bars indicate score variability.

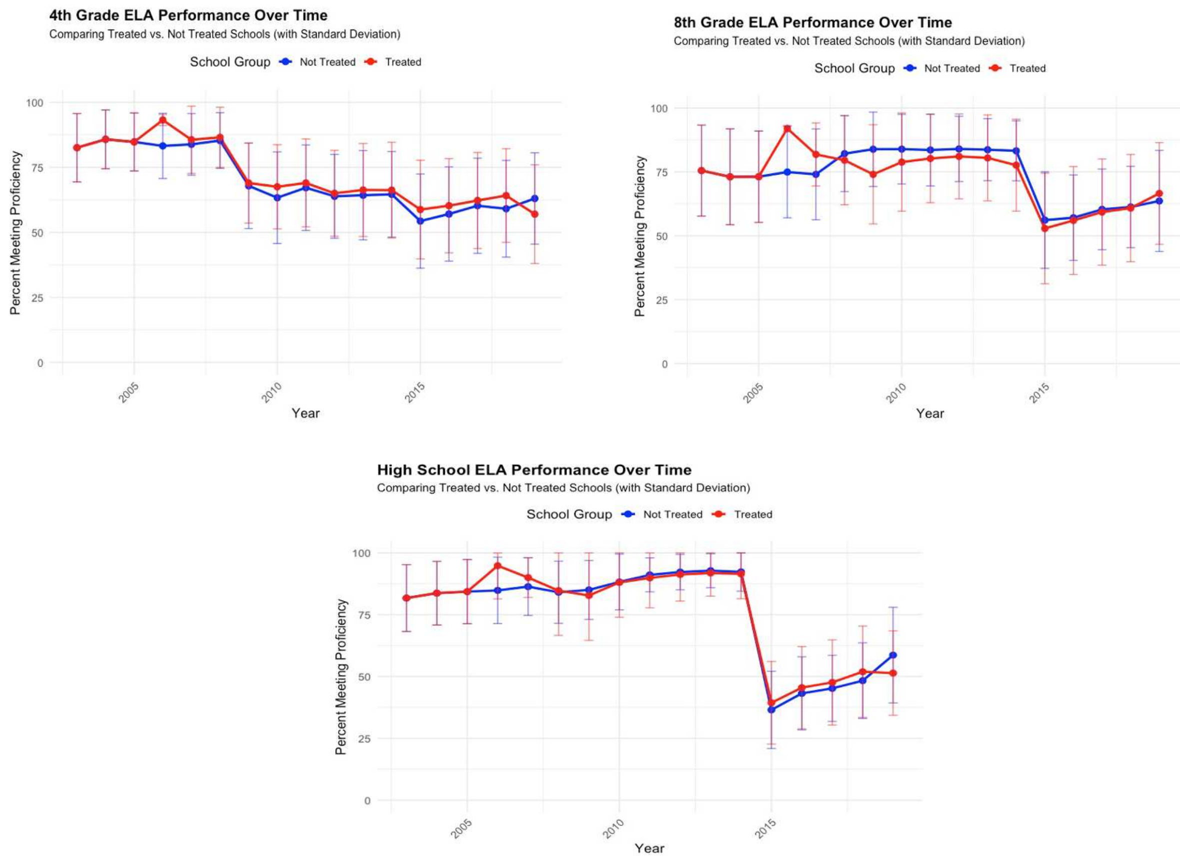


Figure A2: Evolution of ELA Scores by Grade level

Note: This figure presents the average ELA scores for treated and not treated schools over time. “Not Treated" includes both schools that have never implemented the minimum salary policy and those that had not yet done so in a given year. “Treated" schools are those that had already implemented the policy. Because treatment adoption is staggered over time, simple mean comparisons do not denote causal effects. Standard deviation bars indicate score variability.

Table A3: Difference-in-Difference Estimates (Event Study)

Variable	Coefficient	Std. Err.	Lower Bound	Upper Bound
Pre Policy	-322.85	173.92	-663.74	18.03
Post Policy	1234.18	470.56	311.91	2156.46
tm4	-623.83	210.80	-1036.99	-210.68
tm3	-282.90	210.23	-694.95	129.15
tm2	-38.20	218.41	-466.27	389.87
tm1	-346.47	188.86	-716.63	23.70
tp1	474.83	150.16	180.52	769.14
tp2	672.47	268.66	145.91	1199.03
tp3	909.39	394.03	137.10	1681.68
tp4	998.36	530.15	-40.71	2037.44
tp5	1622.20	546.57	550.95	2693.45
tp6	1564.13	586.87	413.89	2714.37
tp7	1777.37	695.86	413.51	3141.23
tp8	1854.71	934.52	23.08	3686.34

Note: Event Study estimates of the increase in teacher salaries. tp1 denotes the increase in salary one year after the negotiation passed. Post policy averages the values upto eight years after treatment.

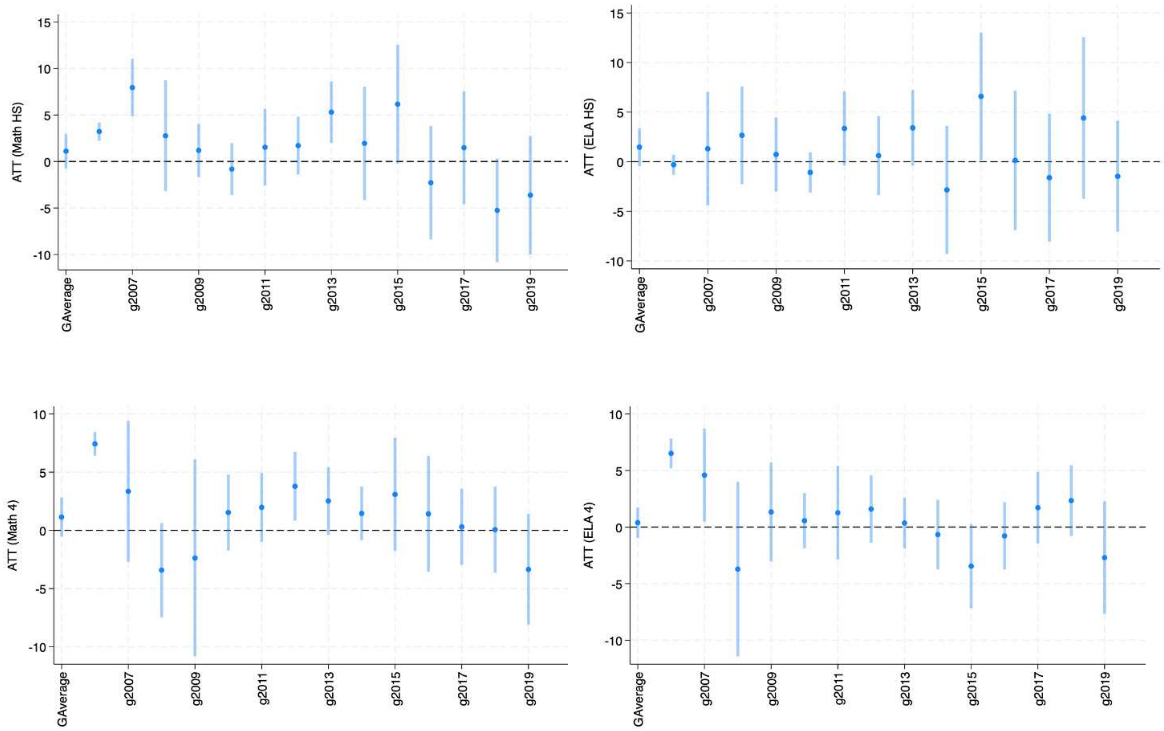


Figure A3: Treatment Effects by Group

Note: This figure presents the estimated Average Treatment Effects on the Treated (ATT) for high school and 4th-grade Math and ELA scores, disaggregated by the year in which schools received the salary increase. The “GAverage" represents the pooled effect across all treated schools. Treatment here is binary instead of continuous treatment to ensure better comparability across years.

Table A4: Estimation of Treatment Effects: Average Total Effect

Score	Estimate	SE	N	Switchers	Estimate/SD	Estimate/Mean
Math 4	1.17	0.45	10,305	4,231	0.05	0.017
Math 8	-0.0785	0.625	6,000	2,293	< 0.01	< 0.01
HS Math	1.72	0.49	3,606	1,459	0.06	0.026
ELA 4	0.73	0.39	10,339	4,255	0.03	0.010
ELA 8	0.37	0.51	6,911	2,746	0.02	0.001
ELA HS	1.48	0.56	3,767	1,596	0.06	0.010
Grad. Rate	0.93	0.31	3,807	1,629	0.07	0.005
College Grad. Rate	0.44	0.41	2,850	982	0.04	0.001

Note: Table reports the average treatment effect of implementing the new salary guide on student outcomes. N represents the total number of observations (unbalanced). The fourth column reports the total number of schools that switched treatment across the time period. The fifth column shows the ratio of estimate to standard deviation, and the last column shows the ratio of estimate to mean scores.

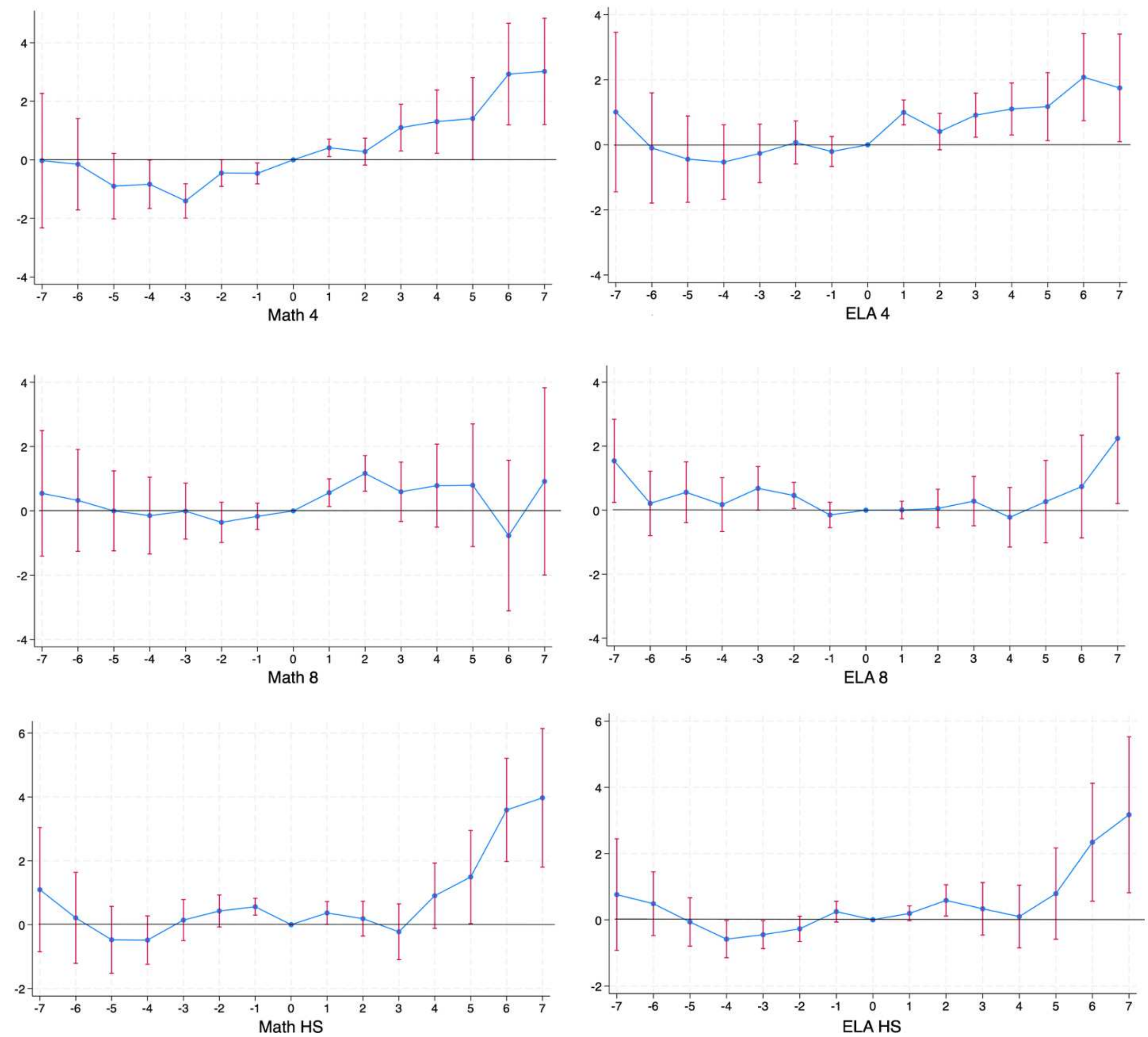


Figure A4: Same Switchers (Y-axis is change in Percentage of students meeting proficiency)

Note: Point estimates and 95% confidence interval of parameters β_k in Equation 1.2 at the school level constructed using valid Difference-in-Differences (DiD) comparisons between treated districts and not-yet-treated districts.. Y-axis shows changes in percentage points of students meeting proficiency. X-axis refers to time period before and after the treatment. Reference year is time period 0, which indicates the year before treatment took place. Treatment is non-binary and run on a strongly balanced panel(same-switchers). Although not shown here, effects hold for graduation rates as well. I cannot run a same switcher model estimate on percentage of students enrolled in college due to missing data for years 2011 and 2012.

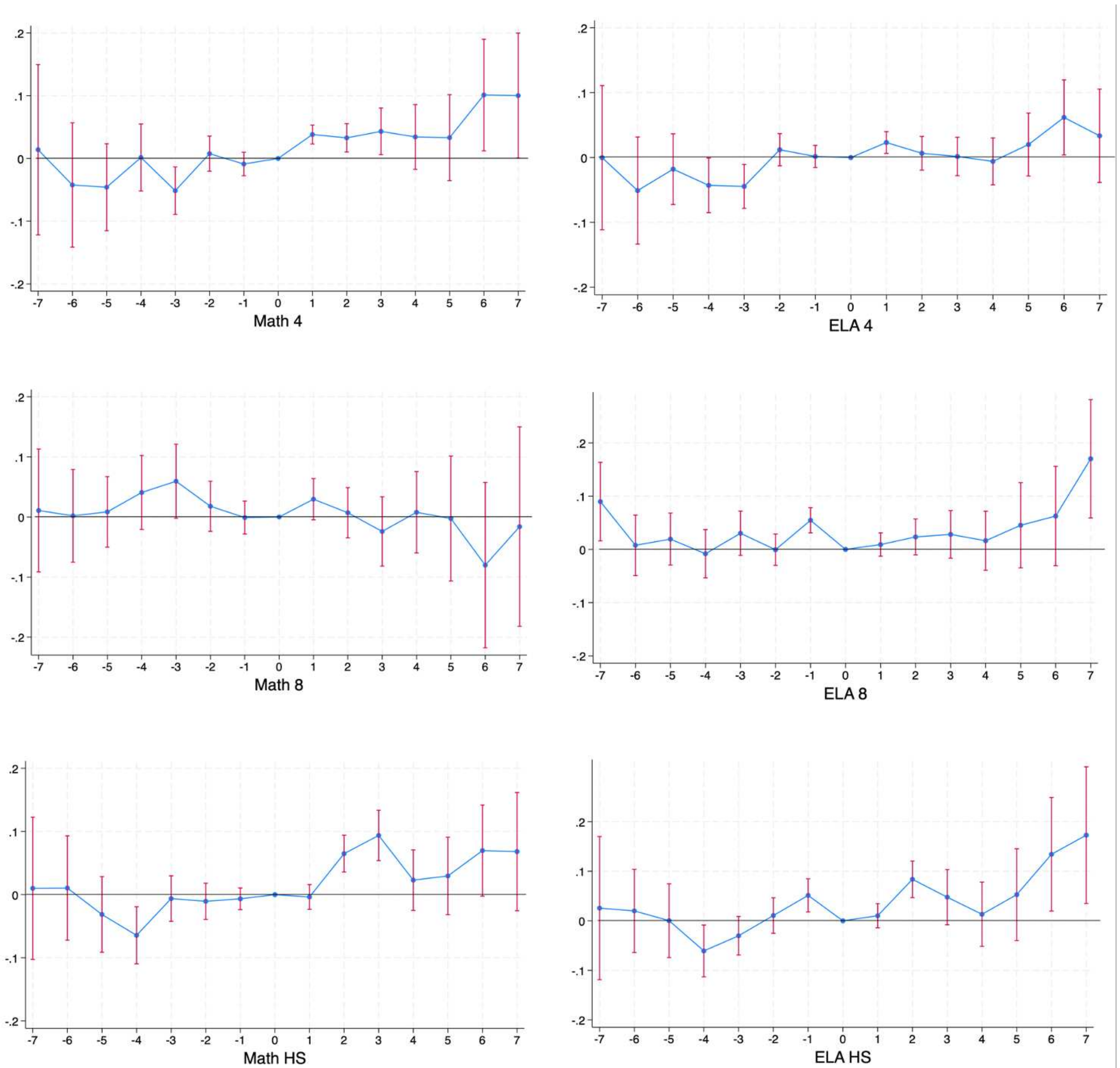


Figure A5: Standardized Scores

Note: Point estimates and 95% confidence interval of parameters β_k in Equation 1.2 at the school level constructed using valid Difference-in-Differences (DiD) comparisons between treated districts and not-yet-treated districts.. Y-axis shows changes in standardized values of percentage of students meeting proficiency. Each outcome is standardized by year. X-axis refers to time period before and after the treatment.

Event Study using Alternate Estimators

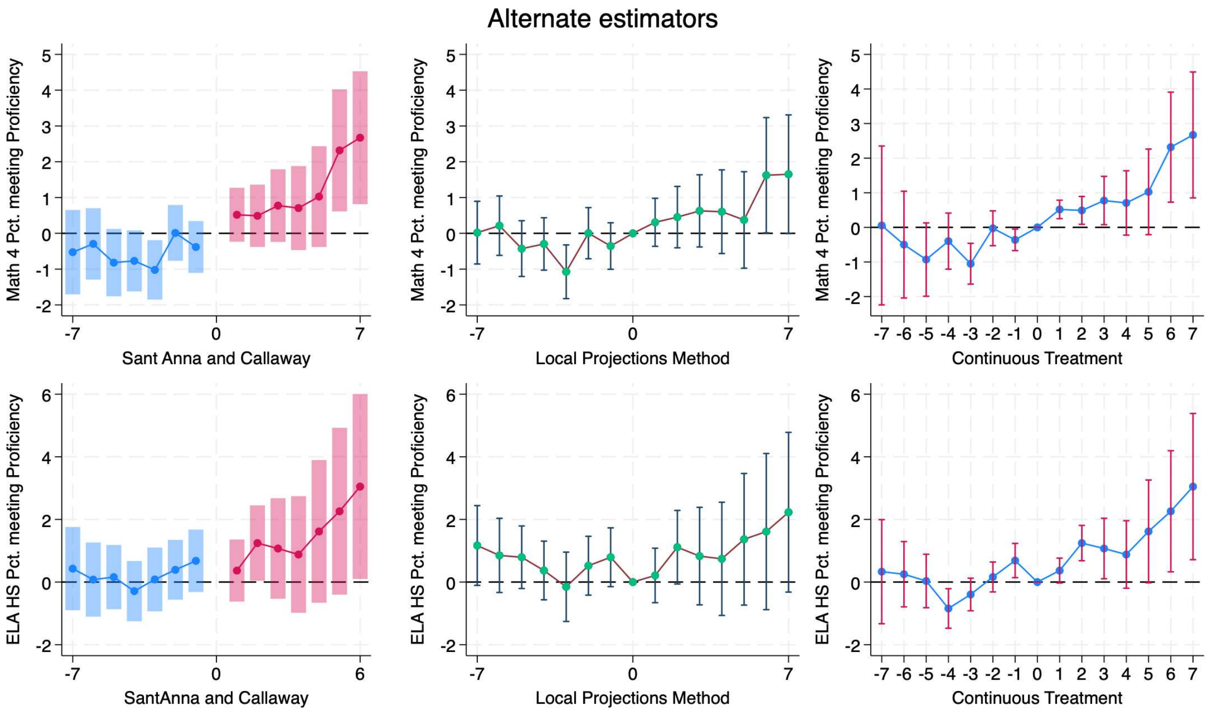


Figure A6: Alternate Estimators

Note: Point estimates and 95% confidence interval of parameters β_k in Equation (1.2) constructed using valid Difference-in-Differences (DiD) comparisons between treated districts and not-yet-treated districts.. The first columns shows event study using Callaway and SantAnna DiD estimators. Pink shaded areas denote post treatment periods and blue shaded areas denote pre treatment periods. The white gap is the reference period. The second column shows event study using the Local projections DiD. The third column showns event study using continuous treatment. This is the same event study I show in the results of the paper. Included here for reference. Although I only show results for Math 4 and high school ELA, results hold for all other grades suggesting that the results are robust to using alternate estimators. Standard errors are the largest when using SantAnna and Callaway.

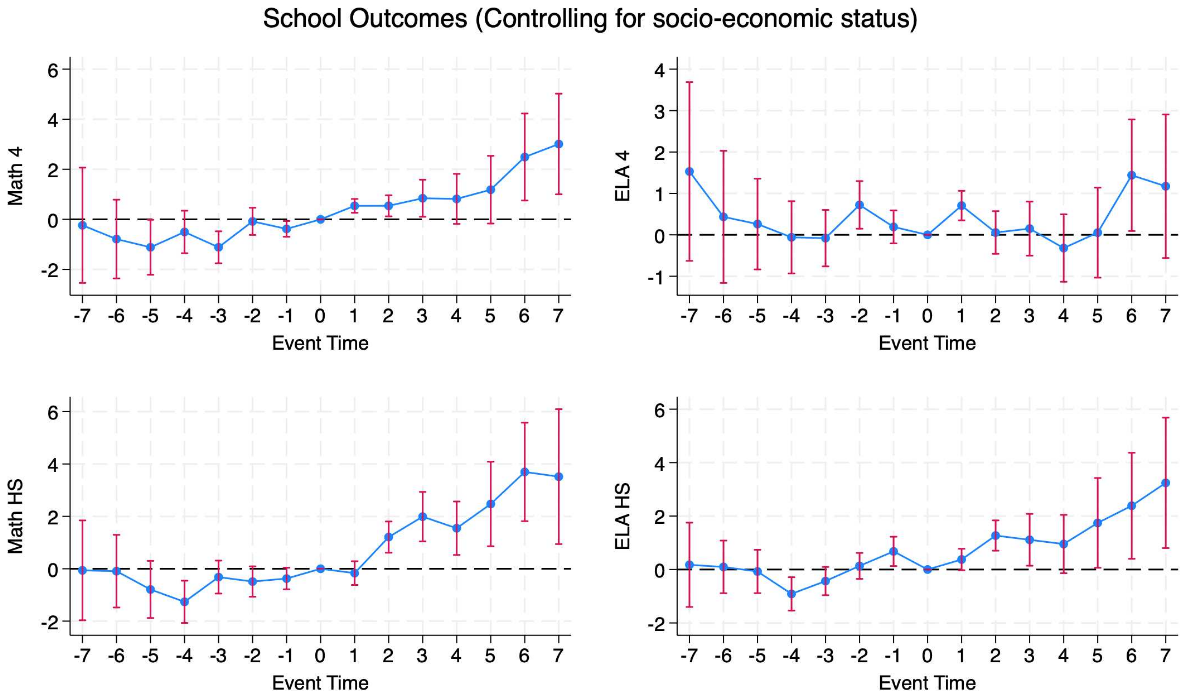


Figure A7: Event Study by DFG

Note: Point estimates and 95% confidence interval of parameters β_t in Equation (1.2) controlling for socio-economic status based on district factor group, and constructed using valid Difference-in-Differences (DiD) comparisons between treated districts and not-yet-treated districts.. The classification is done based on Table A5.

Table A5: Distribution of Socio-economic Classifications Across Categories

Classification	**A**	**B**	**CD**	**DE**	**FG**	**GH**	**I**	**J**	**Total**
1	3,927	3,230	2,669	0	0	0	0	0	9,826
2	0	0	0	3,927	3,995	4,284	0	0	12,206
3	0	0	0	0	0	0	5,457	1,156	6,613
Total	3,927	3,230	2,669	3,927	3,995	4,284	5,457	1,156	28,645

From lowest socioeconomic status to highest, the categories are A, B, CD, DE, FG, GH, I, and J. The numbers denote total observations across all years for all schools.

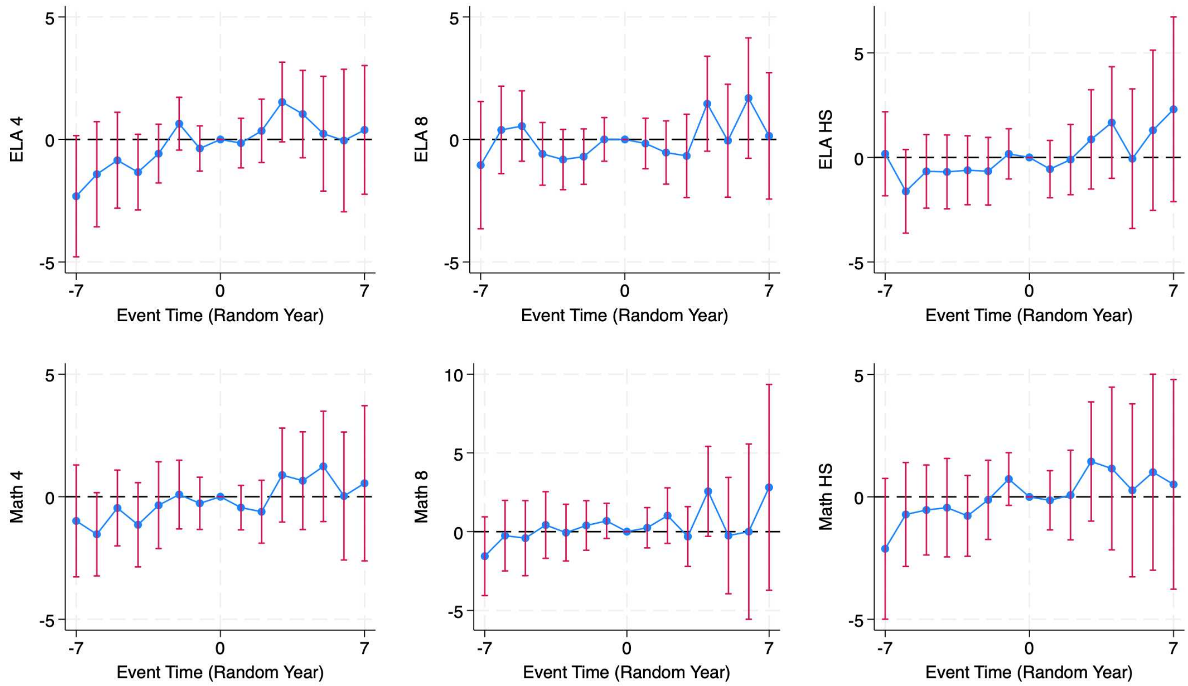
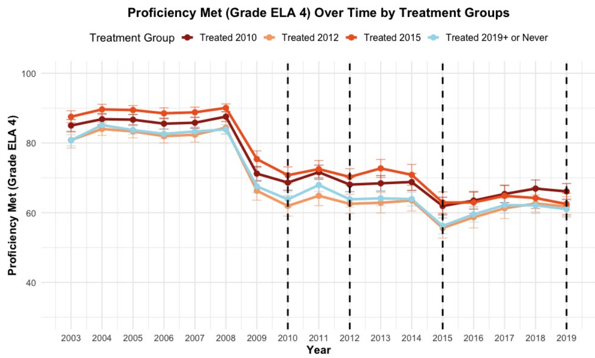
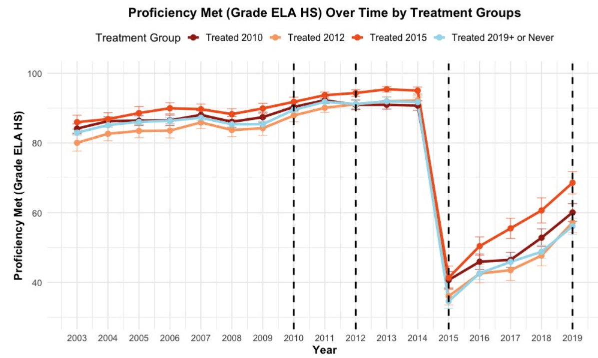


Figure A8: Evolution of Scores: Random Treatment

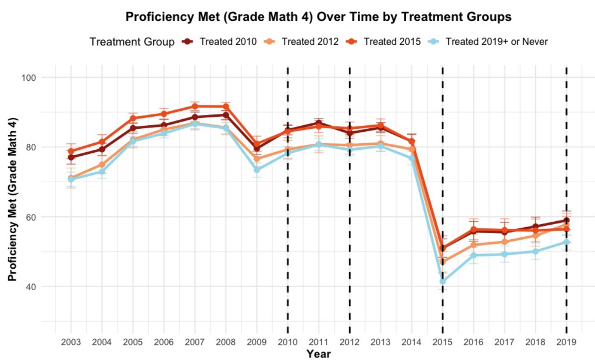
Note: Point estimates and 95% confidence interval of parameters β_k in Equation (1.2) constructed using valid Difference-in-Differences (DiD) comparisons between treated districts and not-yet-treated districts.. Treatment year is set three years from the actual year the negotiation was passed and treatment level is randomized. Event study shows no positive effects when the year of treatment is randomized.



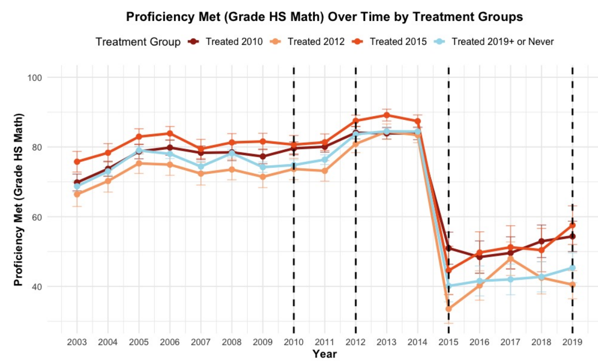
(a) ELA Grade 4



(b) ELA High School



(c) Math Grade 4



(d) Math High School

Figure A9: Evolution of Scores by treatment

Note: The figures present average scores over time for schools treated 2010, 2012, 2015 and 2019 or never. Event study figures plot the differences in expected means whereas these figures show if patterns changed following treatment.

Teacher Mobility and Inequality: Theoretical Model

This model examines how wage differentials between school districts lead to sorting within a fixed pool of teachers, exacerbating inequalities in teacher quality and student outcomes.

Assumptions

- **No significant increase in Teacher Pool:** The total number of teachers N is relatively constant i.e. individuals who did not want to be teachers do not get into teaching due to the increase in pay.

- **Teachers:** Each teacher i has a quality level $q_i \sim F(q)$, where $F(q)$ is the distribution of teacher quality with mean μ_q and variance σ_q^2.

- **Districts:** Each district j offers a wage w_j, and districts differ in student composition, represented by λ_j, the fraction of low-achieving students.

- **Utility Function:** The utility of teacher i in district j is:

$$U_i^j = w_j + \phi(\lambda_j) + \epsilon_i^j$$

  where $\phi(\lambda_j)$ represents non-pecuniary disutility from teaching in low-performing districts, and ϵ_i^j is a teacher-specific preference shock.

Districts' Problem: Setting Wages

Each district j maximizes the average quality of its teachers q_j by offering competitive wages, subject to its budget constraint:

$$\max_{w_j} q_j = E[q_i \mid U_i^j \geq U_i^k \quad \forall k]$$

subject to:

$$w_j N_j \leq B_j$$

where N_j is the number of teachers and B_j is the district's budget.

Teachers' Decision: Sorting Based on Utility

Teachers sort themselves across districts to maximize their utility. Teacher i will choose district j over district k if:

$$U_i^j = w_j + \phi(\lambda_j) \geq w_k + \phi(\lambda_k)$$

Rearranging:

$$w_j - w_k \geq \phi(\lambda_k) - \phi(\lambda_j)$$

This shows that a wage differential $w_j - w_k$ must be large enough to compensate for differences in student composition.

Wage Differentials and Teacher Quality

Given the fixed teacher pool, the average quality of teachers in district j is:

$$q_j = E[q_i \mid U_i^j \geq U_i^k \quad \forall k]$$

High-wage districts attract better teachers, so $q_H > q_L$, where q_H and q_L represent the average quality of teachers in high- and low-wage districts, respectively.

Expanding Inequality (with Fixed Teacher Pool)

Wage differentials lead to growing inequality in teacher quality:

$$\Delta q = q_H - q_L$$

As wage differentials $w_H - w_L$ increase, this gap expands over time:

$$\Delta q_t = \Delta q_{t-1} + f(w_H - w_L)$$

Even though N is fixed, wage competition reallocates teachers, causing the inequality in teacher quality and student outcomes to persist:

$$\lim_{t \to \infty} \Delta q_t > 0$$

Based on this simple theoretical framework, I show that inequality may exacerbate.

Appendix Chapter 2

TSLS Estimates with expanded covariates

Expanded covariates model includes additional variables to create a local window. These variables include entry cohort, gender, early STEM interest, and age at entry. Our results hold under these different set of student samples as well. The inclusion of these covariates leads to a more conservative local window ([13,748, 18,252]), resulting in a smaller effective sample size. Despite this restriction, the confidence interval for the first-year persistence effect remains entirely above zero, ranging from 0.172 to 0.968, thereby reinforcing the robustness of the finding. The interval for second-year persistence is wider and includes zero ([-0.046, 0.786]), implying that while the estimate is suggestive of a positive effect, statistical uncertainty remains and the estimate is significant only at the 90% confidence interval. Similarly, the confidence interval for completed credits is quite wide and spans both negative and positive values ([-1.141, 14.757]), precluding definitive conclusions. Lastly, the estimated impact on GPA remains close to zero and the confidence interval again crosses zero, indicating a lack of detectable effect.

Table A6: TSLS Estimates with Expanded Covariates (EFC Cutoff = $16,000)

Outcome	**TSLS Estimate**	**95% CI**	**Obs (Below / Above)**
Persistence (Complete 1st Yr)	0.570***	[0.172, 0.968]	21 / 19
Complete 2nd Year	0.370*	[-0.046, 0.786]	21 / 19
First-Year Completed Credits	6.808	[-1.141, 14.757]	19 / 16
Standardized GPA (Spring Yr 1)	0.040	[-0.008, 0.088]	21 / 19

Note: Estimates are based on Two-Stage Least Squares (TSLS) under a fuzzy local randomization design centered at an EFC cutoff of $16,000. The selected window is [13748, 18252] with 40 total observations. Covariates include high school GPA, first-year need-based aid, merit scholarships (excluding BSP-specific awards), entry cohort, gender, early STEM interest, and age at entry. Confidence intervals reflect asymptotic normality.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

This section presents a consolidated summary of intention-to-treat (ITT) estimates under the local randomization regression discontinuity design (RDD) framework. Although the main text emphasizes treatment-on-the-treated (TOT) estimation via two-stage least squares (TSLS), ITT effects are policy-relevant and more directly interpretable, particularly given the high compliance rate near the cutoff. Each row in Table A7 represents a separate estimation window or subsample.

Table A7: Consolidated Local Randomization ITT Estimates Across Subsamples and Specifications

Outcome	ITT Estimate	p-value	95% Confidence Interval	Obs Below Cutoff	Obs Above Cutoff
First-Year Persistence	0.335***	0.000	[0.300, 0.660]	92	43
Second-Year Completion	0.215**	0.022	[0.070, 0.520]	92	43
First-Year Completed Credits	1.079	0.444	[-2.210, 4.540]	84	34
STEM Participation (Spring Term)	0.043	0.840	[-0.340, 0.200]	84	34
First-Year Persistence (Expanded Covariates)	0.407**	0.016	[0.250, 0.970]	21	20
Second-Year Completion (Expanded Covariates)	0.264	0.160	[0.030, 0.750]	21	20
First-Year Credits (Expanded Covariates)	4.658	0.120	[-0.840, 12.930]	19	16
First-Year Persistence (Age 19)	0.292**	0.046	[0.180, 0.740]	25	30
Second-Year Completion (Age 19)	0.092	0.724	[-0.190, 0.510]	25	30
First-Year Credits (Age 19)	2.306	0.334	[-2.200, 8.870]	23	24
First-Year Persistence (Wider Window)	0.286***	0.004	[0.270, 0.630]	87	33
Second-Year Completion (Wider Window)	0.171	0.100	[0.090, 0.490]	87	33
Fall-Spring Persistence (Fake Cutoff = $5k)	-0.078	0.472	[-5.280, 4.080]	43	35
Second-Year Persistence (Fake Cutoff = $5k)	0.036	0.872	[-21.600, 15.100]	43	35
Fall-Spring Persistence (Fake Cutoff = $8k)	-0.120	0.160	[-2.820, 0.780]	68	63
Second-Year Persistence (Fake Cutoff = $8k)	-0.069	0.332	[-1.500, 1.100]	68	63

Note: All estimates use local randomization RDD centered at the $16,000 EFC cutoff. Each row corresponds to a separate specification, subsample, or falsification test. Confidence intervals were obtained via randomization inference or large-sample approximation, as appropriate. While not shown for brevity, estimates with lower number of observations mean I included more covariates. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Across specifications, the strongest and most consistent finding is that BSP eligibility significantly improves persistence into the second year of college. Confidence intervals are tight and consistently exclude zero, suggesting a robust, practically meaningful program effect. While estimated effects on second-year completion and credit accumulation are positive, they are less precise and do not reach conventional significance levels in all specifications. STEM participation appears unaffected.

These ITT estimates serve as lower bounds on causal effects for compliers, given the high take-up rate near the cutoff. The consistency across covariate sets, window sizes, and subsamples enhances confidence in the causal interpretation. Moreover, placebo tests using false thresholds provide no evidence of spurious effects, further validating the identification strategy.

Nevertheless, the Fisherian framework's emphasis on confidence intervals over point estimates cautions against overinterpreting any single estimate. The evidence should be read as probabilistic support for the existence of non-zero effects, especially on persistence, rather than precise quantification of the treatment effect magnitude.

Scholarship effects

Merit Based Scholarship

Figure A10 demonstrates the probability of receiving the scholarship (hereby referred to as the GnG scholarship) for students based on their high-school GPA (HSGPA). There is a clear jump in scholarship probability at an HSGPA of 3.9, validating the first stage of our fuzzy RDD. The amount offered ranges from $500-$4000. The average award is $2369 with a standard deviation of $637.

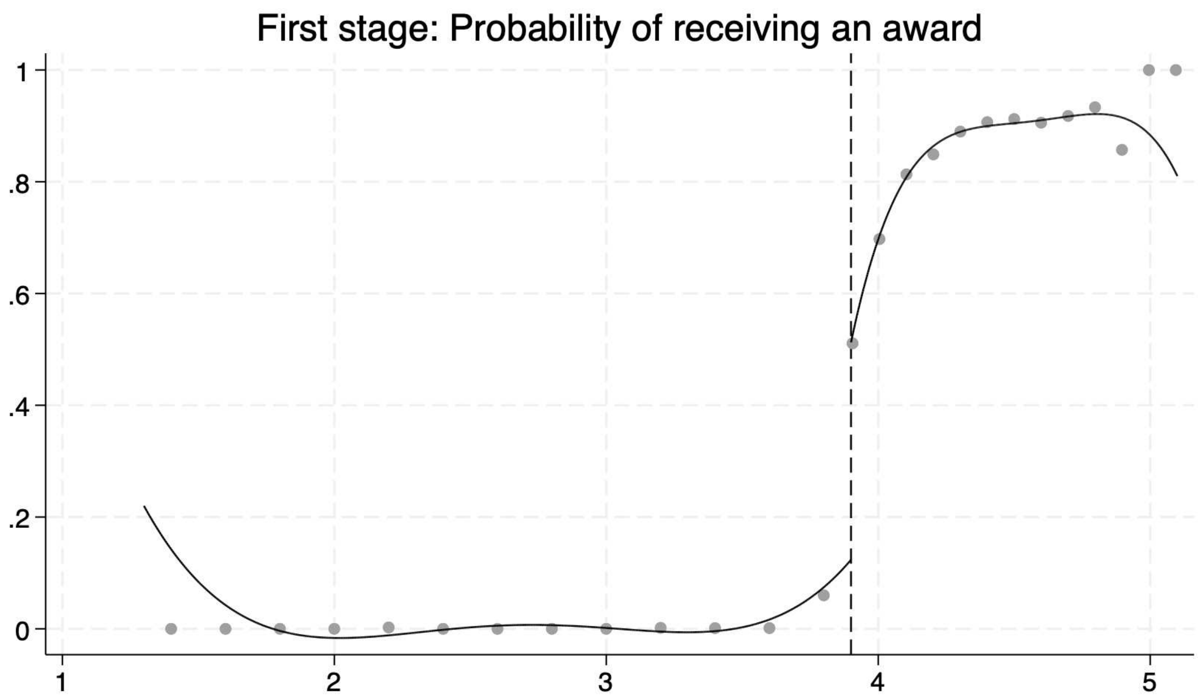


Figure A10: First Stage Regression Discontinuity Plot

Note: This figure shows the probability of receiving the GnG scholarship. The jump at an HSGPA of 3.9 reflects the first stage of the fuzzy RDD.

Results

Tables A8 and A9 present the treatment effect estimates for the pre-COVID and post-COVID periods, respectively. The analysis includes covariates such as race, gender, and first-generation status. Before the pandemic, the GnG scholarship had no significant effect on GPA or STEM enrollment but demonstrated substantial positive effects on persistence outcomes. Scholarship recipients showed a 45-percentage-point increase in persistence rates compared to non-recipients ($\tau = 0.45$), significant at the 1% level. This suggests that the financial aid component helped alleviate barriers to student retention. Similarly, the scholarship increased second-year completion rates by 28 percentage points ($\tau = 0.28$), further highlighting its impact on sustained enrollment.

While STEM enrollment saw a small positive estimate ($\tau = 0.10$), the effects were not statistically significant. This aligns with the scholarship's broader design, which does not explicitly target STEM outcomes.

Table A8: Pre-COVID Treatment Effect Estimates for GnG Scholarship (Covariate-Adjusted Fuzzy RDD)

Outcome	Method	Estimate	P>z	95% Confidence Interval	Eff. Obs.
1st Year GPA	Bias-corrected	0.14	0.29	[-0.12, 0.39]	4121/1320
STEM Enrollment (1st Year)	Bias-corrected	0.10	0.34	[-0.11, 0.31]	4123/1320
Persist to 2nd Year	Bias-corrected	0.45	0.00	[0.33, 0.58]	4451/1358
Complete 2nd Year	Bias-corrected	0.28	0.00	[0.14, 0.42]	4451/1358

Note: Results are derived from a fuzzy RDD analysis of the Green and Gold scholarship before the pandemic.

Post-pandemic, the persistence effects weakened significantly. The impact on second-year persistence dropped to an estimate of $\tau = 0.013$, which was no longer statistically significant. This likely reflects broader structural disruptions caused by the pandemic. Similarly, second-year completion showed weaker effects ($\tau = 0.076$), indicating that the scholarship's earlier advantages in promoting retention did not hold under post-COVID conditions. STEM outcomes remained consistently insignificant, with the robust estimation methods yielding null effects. GPA improvements also remained modest and statistically imprecise.

Table A9: Post-COVID Treatment Effect Estimates for GnG Scholarship (Covariate-Adjusted Fuzzy RDD)

Outcome	**Method**	**Estimate**	**P>z**	**95% Confidence Interval**	**Eff. Obs.**
1st Year GPA	Bias-corrected	0.099	0.45	[-0.16, 0.36]	1215/600
STEM Enrollment (1st Year)	Bias-corrected	0.054	0.59	[-0.14, 0.25]	1216/602
Persist to 2nd Year	Bias-corrected	0.013	0.80	[-0.09, 0.12]	1324/632
Complete 2nd Year	Bias-corrected	0.076	0.24	[-0.05, 0.20]	1324/632

Note: Results are derived from a fuzzy RDD analysis of the Green and Gold scholarship after the pandemic. Post-pandemic effects show weaker impacts of the GnG scholarship, consistent with changes in broader academic and social contexts.

Need-Based Aid

This section examines the potential effects of need-based aid on student outcomes. Need-based aid are aids awarded to students based on their Expected Family Contribution (EFC) and these awards do not need to be repaid. I employ a Sharp Regression Kink Design (RKD) (see Card et al. (2017)) to evaluate the causal impact of need-based financial aid on student outcomes. This method leverages a change in the slope of the treatment assignment function at a specific threshold—in this case, the Expected Family Contribution (EFC) cutoff of $6,000. Figure A11 illustrates the kink in the distribution of need-based aid at the $6,000 EFC threshold, indicating a shift in financial aid policy.

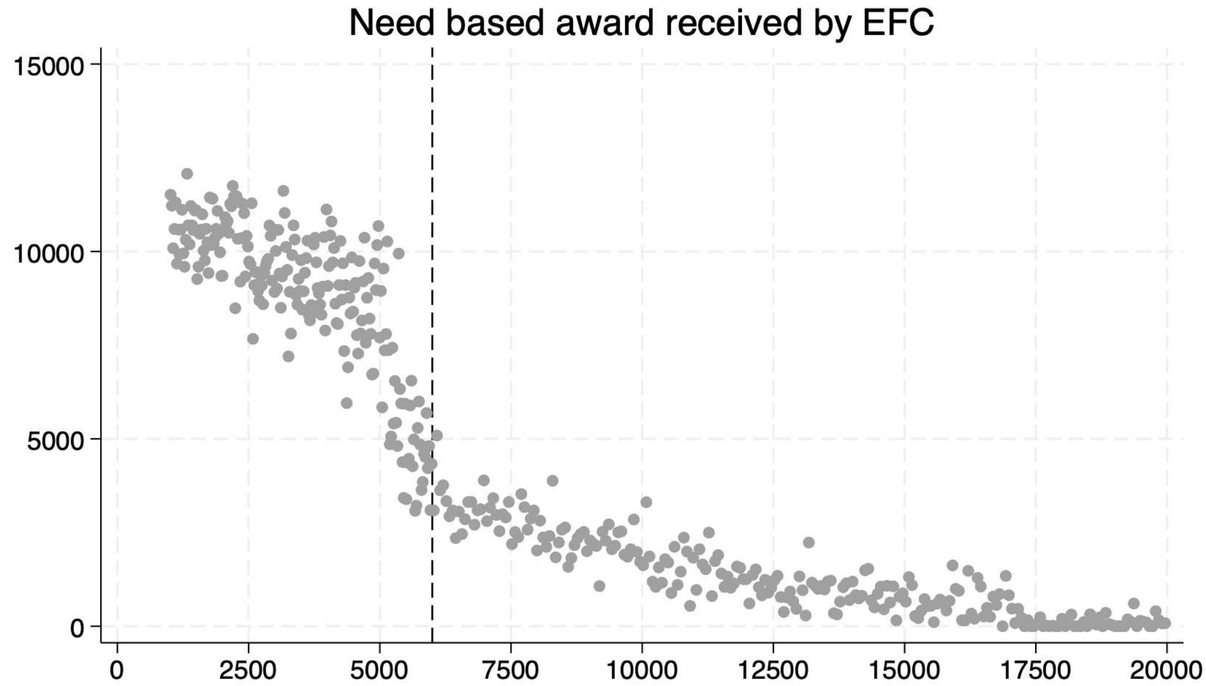


Figure A11: Need-Based Aid by EFC: Evidence of a Slope Change at $6,000

Note: This figure depicts the distribution of need-based aid relative to Expected Family Contribution (EFC). The sharp change in slope at $6,000 reflects the policy threshold.

A crucial assumption for RKD validity is the absence of similar slope changes in other covariates around the threshold. Figures A12, A13, and A14 demonstrate that there are no significant slope changes in merit-based aid, loan allocations, or HSGPA, confirming the validity of our design.

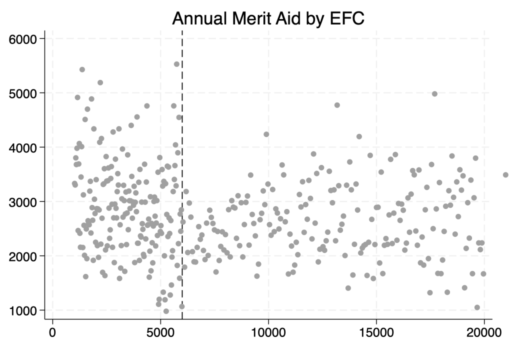


Figure A12: Merit-Based Aid by EFC

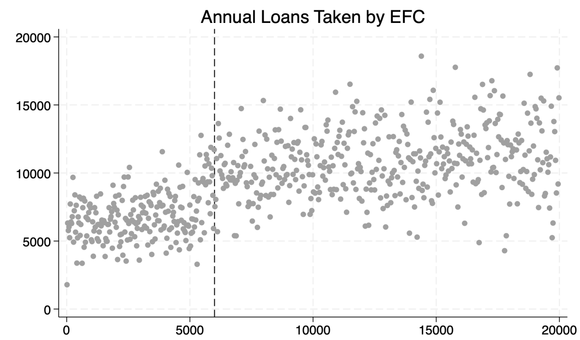


Figure A13: Loan Allocations by EFC

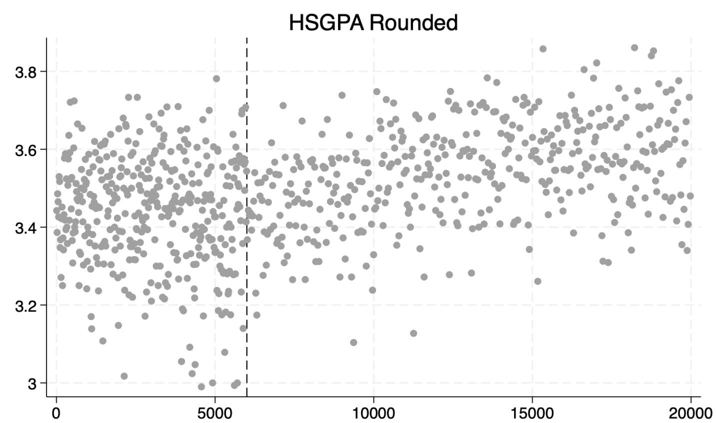


Figure A14: HSGPA by EFC

Note: The figures display the relationship between EFC and three financial and academic outcomes: merit-based aid, loan allocations, and high school GPA. Across all three outcomes, there is no visible slope change at the $6,000 EFC threshold, suggesting that the threshold does not serve as a distinct cut-off point for these measures. The top row presents the results for merit-based aid and loan allocations, while the bottom figure focuses on high school GPA. The analysis uses a fuzzy RKD framework to assess slope continuity at the specified EFC threshold.

Results

Table A10 presents RKD estimates across key academic outcomes. The results suggest that need-based aid at the $6,000 EFC threshold does not meaningfully influence persistence. While the estimate for 1st Yr GPA is statistically significant, the magnitude is minimal. Specifically, for every $1,000 increase in EFC, GPA decreases by 0.058 units, implying that a $2,500 increase in EFC would reduce GPA by 0.125 units given students on the right side of the $6K EFC receive less need-aid.

Table A10: Treatment Effect Estimates Using Covariate-Adjusted Sharp RKD

Outcome	**Method**	**Estimate**	**P>z**	**95% Confidence Interval**	**Eff. Obs.**
1st Yr. GPA	Bias-corrected	-0.000058	0.023	[-0.00011, -0.000008]	2212/1614
2nd Yr. GPA	Bias-corrected	-0.000040	0.067	[-0.000082, 0.000003]	1974/1438
STEM 1st Yr.	Bias-corrected	0.000004	0.719	[-0.000019, 0.000028]	2879/2054
STEM 2nd Yr.	Bias-corrected	0.000007	0.601	[-0.000018, 0.000031]	2478/1786
Persist to 2nd Yr	Bias-corrected	0.000006	0.546	[-0.000014, 0.000027]	2752/1986
Complete 2nd Yr.	Bias-corrected	-0.000002	0.875	[-0.000028, 0.000024]	2475/1832

Note: Estimates are derived from a Sharp RKD using covariate-adjusted models. Eff. Obs. denotes the effective number of observations on each side of the cutoff.

Similarly, the estimate for completing the second year is statistically insignificant. If the upper bound of -0.000028 were the true effect, a $2,500 increase in EFC would reduce the probability of completing the second year by 0.07 units- a 7% point decrease. Overall, given the direction of effect is not consistent between completing first year (persistence to 2nd year) and completing second year(complete 2nd year), increased need aid does not seem to be affecting persistence for these group of students.

Alliance Award Aid on Student Outcomes

This section evaluates the impact of the Alliance Award, a scholarship program at the college designed to support students from 10 select high schools that were selected based on a combination of high rates of free/reduced lunch eligibility, high first-generation college-going rates, and regional representativeness across the state. Students from Alliance schools are among the most structurally disadvantaged and academically underprepared at the college. Initially launched in 2006, the award amount increased from $2,500 to $4,000 annually in

2016. This policy change serves as the basis for a Difference-in-Differences (DiD) analysis to estimate the causal impact of the increased scholarship amount on academic outcomes, using 2015 as the control year. Pre-treatment trends were verified for parallelism (not shown for brevity), confirming the validity of the DiD framework.

The DiD model estimates the outcome Y_{it} for student i who first enrolled in the college at time t:

$$Y_{it} = \alpha + \beta_1 D_i + \beta_2 \text{Post}_t + \beta_3 (D_i \times \text{Post}_t) + \epsilon_{it},$$

where:

- D_i is 1 if the student is from an Alliance high school (treatment group) and 0 otherwise (control group).
- Post_t is 1 for years after 2016 and 0 otherwise.
- β_3, the interaction term coefficient, represents the treatment effect of the increased award on that cohort group.

Results

Table A11 summarizes the DiD estimates for each group of students based on when they entered college. Years 2021–2023 are included but show incomplete data due to recent cohorts still progressing through their programs. Year denotes the college outcomes for groups of students that enter the student that year i.e. year 2017 shows the outcomes of students that entered the college the Fall of 2017 and the estimates are compared to those of students that attended the college the Fall .

The increase in aid shows shows statistically significant positive effects on persistence for students starting college 2016 and 2017 However, effects diminish over time, becoming statistically insignificant by 2019. These findings suggest that the Alliance Award did not have similar effects on groups of students that attended the college at a later date. First-year and final GPA estimates are positive but imprecise. Confidence intervals from 2018 onward

Table A11: Difference-in-Differences Estimates for Alliance Aid Outcomes (2015 Control Year)

Year	1st Yr. GPA	Final GPA	Persist 1st Yr	Persist 2nd Yr	STEM Enrollment	STEM Graduation
2016	0.11 [0.02, 0.20]	0.09 [0.00, 0.18]	0.12 [0.03, 0.21]	0.10 [0.01, 0.19]	0.02 [-0.03, 0.07]	0.04 [-0.05, 0.13]
2017	0.10 [0.01, 0.19]	0.08 [-0.01, 0.17]	0.11 [0.02, 0.20]	0.09 [0.00, 0.18]	0.03 [-0.02, 0.08]	0.03 [-0.06, 0.12]
2018	0.08 [-0.01, 0.17]	0.07 [-0.02, 0.16]	0.09 [0.00, 0.18]	0.08 [-0.01, 0.17]	0.01 [-0.04, 0.06]	0.02 [-0.07, 0.11]
2019	0.07 [-0.02, 0.16]	0.06 [-0.03, 0.15]	0.08 [-0.01, 0.17]	0.07 [-0.02, 0.16]	0.00 [-0.05, 0.05]	0.01 [-0.08, 0.10]
2020	0.06 [-0.03, 0.15]	0.05 [-0.04, 0.14]	0.07 [-0.02, 0.16]	0.06 [-0.03, 0.15]	-0.01 [-0.06, 0.04]	0.00 [-0.09, 0.09]
2021	0.05 [-0.04, 0.14]	NA	0.06 [-0.03, 0.15]	0.05 [-0.04, 0.14]	-0.02 [-0.07, 0.03]	NA
2022	0.04 [-0.05, 0.13]	NA	0.05 [-0.04, 0.14]	0.04 [-0.05, 0.13]	-0.03 [-0.08, 0.02]	NA
2023	0.03 [-0.06, 0.12]	NA	0.04 [-0.05, 0.13]	0.03 [-0.06, 0.12]	-0.04 [-0.09, 0.01]	NA

Note: Confidence intervals at the 95% level. Results for 2021–2023 are incomplete due to ongoing data collection for these cohorts.

overlap zero, indicating these although effects are not statistically significant, I cannot argue effects are null. I also cannot rule out null effects on STEM enrollment.

Taking stock

The analysis in this section assessed the impact of small-scale, institution-specific scholarships on student persistence, academic performance, and STEM enrollment. The findings suggest that while the Green and Gold (GnG) merit-based scholarship significantly increased first-year and second-year persistence for pre-COVID cohorts, it did not produce meaningful effects on cumulative GPA or STEM enrollment. Similarly, the analysis of need-based aid and the Alliance Award yielded mixed results, with no clear evidence that these scholarships substantively improved persistence outcomes.

It is important to emphasize that the absence of statistically significant findings does not imply the absence of effects. Rather, the current estimates are inconclusive, potentially due to data limitations and sample constraints. While the findings do not definitively rule out the possibility that small-scale scholarships influence persistence, the evidence across all three scholarship programs suggests that scholarships alone are unlikely to account for the observed outcomes among Bridge students. The differential impacts of scholarships appear to vary by the student group each scholarship serves, indicating potential heterogeneity in treatment effects. Nevertheless, the magnitude of scholarship effects on persistence remains substantially lower than those observed for the Bridge program, suggesting that scholarships alone are unlikely to be the primary driver of the observed treatment effects. This highlights the need for further investigation into the specific mechanisms through which programmatic supports influence student outcomes beyond financial aid alone.

Appendix Chapter 3

Predicting Baseline Academic Characteristics for Control Selection

In this appendix, I provide supporting evidence for the selection of control variables used in the main analysis of Chapter 3. Because pre-treatment covariates that strongly predict students' academic trajectories are critical for isolating causal effects, I conducted preliminary predictive exercises to assess which student characteristics are most closely associated with STEM enrollment and persistence. This exercise helps inform which variables should be prioritized in the regression adjustments and robustness checks.

Specifically, I used a Random Forest classification model, a machine learning technique well-suited for uncovering nonlinear and interaction-driven relationships, to identify key predictors of (1) enrolling in a STEM major upon college entry and (2) persisting in a STEM major through the beginning of the fourth academic year. Random Forest models aggregate predictions from many decision trees trained on bootstrapped subsamples of the data, reducing overfitting and improving predictive performance (Breiman (2001); Liaw and Wiener (2002)).

To identify factors associated with enrolling as a STEM major during the first semester, I estimated a Random Forest model using pre-college academic indicators: high school GPA (HSGPA), SAT Math scores, and SAT Reading scores.[31]

The model yielded the following variable importances:

- SAT Math: 0.3708
- High School GPA: 0.3740
- SAT Reading: 0.2551

Model predictive performance was reasonable:

[31] ACT scores were also available but omitted from the main prediction exercises due to a lower number of observations. SAT scores provided broader coverage across the sample.

- Test Set Accuracy: 62.85%

- Training Set Accuracy: 69.19%

Figure A15 presents a visualization of the decision tree structure (depth = 2) summarizing the main predictive splits.

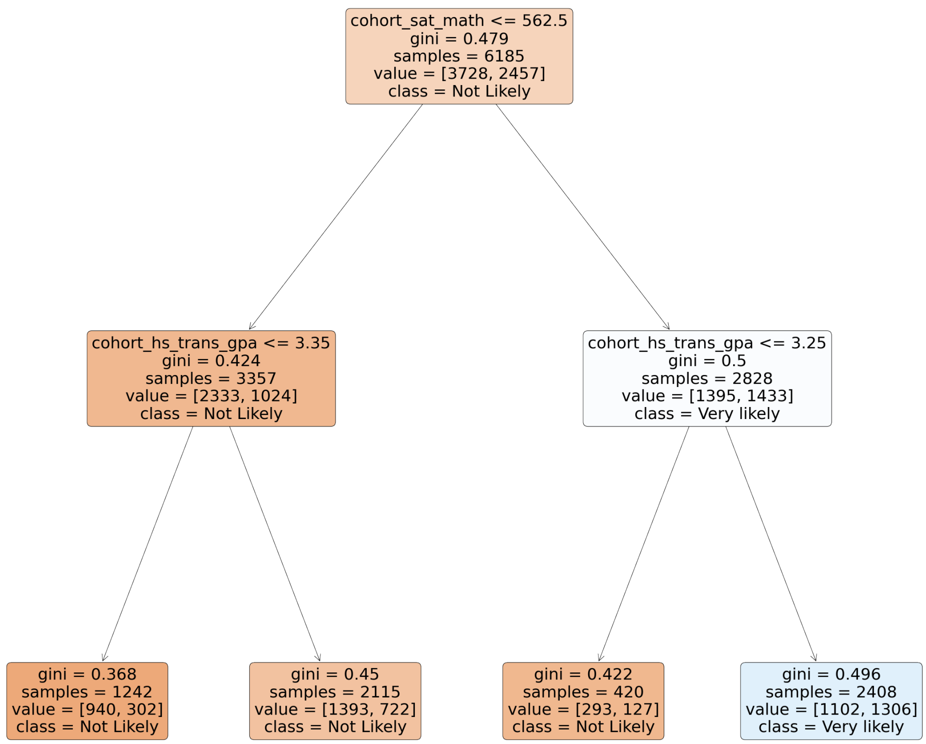


Figure A15: Decision Tree for Initial STEM Enrollment

Interpretation: Students who scored higher than 562.5 on SAT Math and had HSGPA above 3.25 were substantially more likely to enroll in a STEM major during their first semester.

Expanding the feature set to include early college performance indicators (semester GPAs, merit scholarships, high school rank), a second Random Forest model was estimated to predict whether students remained in STEM majors through the beginning of their fourth year.

Variable importance for this model were as follows:

- High School Rank: 0.1386
- End of 3rd Semester GPA: 0.1385
- End of 4th Semester GPA: 0.1378
- End of First Year GPA: 0.1376
- SAT Math: 0.1185
- High School GPA: 0.1100
- SAT Reading: 0.0923
- Gender: 0.0361
- First-Generation Status: 0.0202
- Merit-Based Scholarship Amounts: 0.0199

Model predictive performance improved slightly with a test Set Accuracy of 67.44%.

The corresponding decision tree visualization is shown in Figure A16.

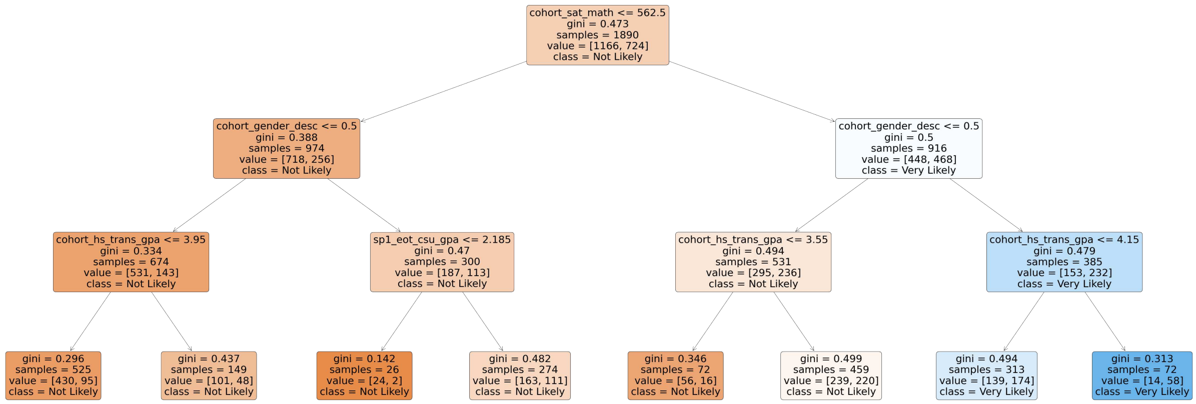


Figure A16: Decision Tree for Fourth-Year STEM Persistence

Interpretation: High SAT Math scores and strong high school academic indicators remained important predictors of long-term STEM persistence. College GPA metrics, unsurprisingly, played a dominant role as students progressed beyond initial semesters.

Based on these predictive results, the following control variables were selected for inclusion in the main Difference-in-Differences specifications:

- Pre-college academic indicators: High School GPA and standardized scores
- College academic performance: Semester-by-semester GPA outcomes.
- Financial variables: Merit-based and need-based scholarship amounts.

These controls capture the major observable factors most predictive of students' major choice trajectories and thus strengthen the empirical validity of the main causal analysis. Including these controls does not materially alter the event-study estimates, with effect sizes remaining consistent with those discussed in the main sections of the paper.