Peer Institution Networks, Test-Optional Admission Policies, and STEM Major Completions

Ethan N. Lewis*

December 2025

Abstract

To investigate the heterogeneous impacts of test-optional policies, I construct a network of private colleges based on their self-reported peer institutions and detect communities within it using a network science algorithm. I then estimate treatment effects by comparing adopters and non-adopters within the same community. Overall, adoption increases *reported* SAT scores but decreases the share of bachelor's degree completions in STEM, consistent with mismatch. Effects vary by community; at the most selective liberal arts colleges, policy adoption causes an 8 (23) percent decline in the share of all (under-represented minority) bachelor's completions with a STEM major.

JEL Codes: I23, I24, J24, C65

*PhD Candidate, Department of Economics, Boston University; email: ethannl@bu.edu. I am grateful to Joshua Goodman, Kevin Lang, and Daniele Paserman for their advice and feedback, as well as participants at the Boston University Microeconomics Dissertation Workshop and the Wheelock College of Education's PREREQ Seminar for helpful comments and questions. I would also like to thank Boston University's Institute for Economic Development for providing financial assistance. Views are my own.

1 Introduction

In this paper, I employ a network-based strategy to identify distinct communities of higher education institutions that can facilitate investigation of heterogeneous treatment effects. Relying on their self-reported peer institutions, I construct a novel network of private colleges and universities and detect communities within it using an algorithm from the network science literature. Communities differ substantially across observables and exhibit different trends in key variables. These distinct trends can even complicate estimation of population-level treatment effects if not accounted for.

I use several of these communities to investigate both the population-level and heterogeneous impacts of adopting a test-optional admission policy[1] on institution-level outcomes. Methodologically, I estimate dynamic, population-level treatment effects by regressing outcomes on treatment leads and lags, as well as institution and *community-time* fixed effects[2]. I estimate community-level treatment effects by regressing outcomes on a post-treatment indicator interacted with dummy variables for community membership (plus the same set of fixed effects). If a parallel counterfactual trends assumption holds conditional on community membership, these regressions consistently estimate population and community average treatment effects on the treated, respectively. For the population-level dynamic estimates, I show how including community-time fixed effects drastically attenuates the pre-trends that can arise when time fixed effects are estimated at the population level.

I document that at 66 private colleges and universities that adopted a test-optional policy between 2006 and 2016, nearly 25 percent of students enrolled in test-optional cohorts did not submit a test score. However, these effects differ substantially across communities; at treated New England Liberal Arts Colleges, nearly one in three enrolled students did not submit an SAT or ACT score. On the other hand, non-submitters make up just 7 and 11 percent, respectively, of test-optional cohorts at Elite National Universities and Liberal Arts Colleges

[1]These policies give applicants the option not to submit a college entrance exam (SAT or ACT) score when applying for admission. Evaluating institutions claim that the decision not to submit a test score is not viewed negatively (Syverson et al. (2018)). For an excellent review of the origins and rise of test-optional admission policies, see Furuta (2017).

[2]Because this setting features staggered treatment timing and effects may be heterogeneous and/or delayed, I estimate treatment effects with the two-stage difference-in-differences estimator of Gardner et al. (2024).

in the Midwest, and treated members of Less Selective National Universities seem to enroll virtually no non-submitters. I find that test-optional policy adoption substantially increases the 25^{th} and 75^{th} percentiles of the reported SAT score distribution. While dynamic effects are modest at treatment onset, they increase thereafter, with the largest effects coming in the third post-treatment period, where point estimates are 22 and 15 points, respectively, and significant at the 1% level. Increases are most significant for the community of Liberal Arts Colleges in New England and the community of Colleges and Universities on the East Coast, who also enroll the largest share of non-submitting students. These results are consistent with a censoring of the left tail of the reported score distribution. In an appendix, I investigate outcomes related to student-body diversity. My results on population-level effects largely agree with Bennett (2022) that test-optional policies increase racial and socioeconomic diversity; however, community-level impacts are imprecisely measured, and I cannot establish meaningful cross-community heterogeneity.

Motivated by the literature on ability sorting across majors, I create and investigate a new outcome: the share of bachelor's degree completions with a STEM major (overall and by under-represented minority (URM) status), offering new evidence on how test-optional policies shape academic choices and downstream outcomes. Across the full sample, policy adoption causes statistically significant declines of around 5 percent in the overall share of graduates with a STEM major, while effects on URM graduates are indistinguishable from zero. I show in an appendix that when counterfactual trends are estimated at the population-level, a negative pre-trend emerges, complicating causal inference; when I construct an alternative grouping based on Carnegie Classifications, I find a null effect. Thus, my network-based identification strategy leads to meaningfully different results than alternative approaches.

I also find substantial cross-community heterogeneity in the impact of test-optional policies on STEM completions. The community of Colleges and Universities on the East Coast and the community of Elite National Liberal Arts Colleges see the most precisely estimated declines in both the overall and URM-only share of graduates majoring in a STEM field. In those communities, adoption causes a 16 and 8 (26 and 23) percent decline, respectively, in the share of all (URM) graduates majoring in a STEM field. All four estimates are statistically significant, and three are significant at the 1% level. The other communities see null effects; accordingly, I can also reject at the 1% level that effect sizes are the same across communities.

The decline in STEM graduates at policy adopters could simply be driven by students who did not submit a test score (non-submitters) preferring different majors; however, it is also consistent with mismatch. Theory and evidence (see Dessein et al. (2025) and Kelly (2022)) suggest that non-submitters are substantially less academically prepared than their submitting peers, which may cause them to switch out of STEM majors at a higher rate or not pursue them in the first place (despite wanting to). To the extent that this mismatch is driving the estimated decline in STEM bachelor's completions, test-optional policies may be adversely affecting students who would have completed a STEM degree at an institution for which they are better academically matched. However, determining the relative importance of these two mechanisms would require micro-level data, and therefore is an avenue for future research.

To my knowledge, this is the first paper that uses network data and community detection in a policy evaluation setting; however, this approach has applicability beyond the higher education context. It would be particularly relevant in settings where access to high-dimensional covariates is limited – precluding the use of other matching techniques – but units can be linked in a network structure. These include development contexts with social network information, as well as settings with private firms; in the latter, connections between firms could be determined by worker flows.

This paper contributes to the growing literature investigating the effects of test-optional admission policies with institution-level data[3]. Early studies focused on selective liberal arts colleges only (Belasco et al. (2015); Rosinger and Ford (2019)), finding little effect on racial or socioeconomic diversity, but positive effects on reported SAT scores. Using two-way fixed effects to estimate treatment effects, Saboe and Terrizzi (2019) focus on a broader set of adopting institutions; they similarly find no effect on student-body diversity, but in contrast find imprecisely estimated declines in reported SAT scores. However, the control institutions

[3]Several studies investigate test-optional policies with admissions data (see Robinson and Monks (2005), Conlin et al. (2013), Conlin and Dickert-Conlin (2017), Kelly (2022), McManus et al. (2023), and Sirolly et al. (2024).

considered differed substantially from the treated institutions, making it unlikely that the parallel counterfactual trends condition was satisfied.

Like Saboe and Terrizzi (2019), Bennett (2022) and Bevers and Mulholland (2022) consider a broad set of adopters, but consider a more representative set of control institutions; Bennett (2022) finds that test-optional policies lead to a 10 to 12 percent increase in under-represented minority enrollment, a 6 to 8 percent increase in female enrollment, and a 3 to 4 percent increase in Pell grant recipient enrollment. Bevers and Mulholland (2022) report higher average debt burdens among graduates; however, neither re-investigates the impact on reported SAT scores. While my study is similar to both, I make a novel methodological contribution by using peer networks and community detection to identify control units and examine cross-community treatment effect heterogeneity, and a novel substantive contribution by investigating the effect of test-optional policies on STEM major completions. Bennett (2022) compares effects for liberal arts colleges and larger universities, as well as more and less selective institutions; however, the community-level analysis I conduct is much more granular. I also revisit the effect of test-optional policies on reported SAT scores.

The test-optional literature is related to a broader literature examining the relationship between barriers to college entrance exam taking and access to higher education. Studies have investigated state-mandated college entrance exam taking policies (see Klasik (2013), Hurwitz et al. (2015), Goodman (2016), and Hyman (2017)), finding positive effects on 4-year college enrollment rates, with increases especially concentrated among students who would not otherwise have taken a college entrance exam. Financial barriers may also be important; Hurwitz et al. (2017) find that increasing the number of free score sends available to applicants has a positive impact on college attendance and completion. Goodman et al. (2020) find that as-if random SAT retakes increase scores, leading to higher 4-year college enrollment rates, particularly among under-represented students. These studies suggest that reducing barriers to taking and reporting college entrance exams can improve access to higher education; test-optional policies, in contrast, seek to improve access by eliminating the requirement altogether, which may lead to adverse outcomes such as mismatch.

2 Empirical Motivation

To many, the decision to adopt such a policy is puzzling: how could a college benefit from allowing applicants to withhold information? A simple explanation is that colleges' objective functions contain inputs other than their students' academic preparedness. They likely care about the non-academic composition of their student bodies, aiming to enroll legacy students, under-represented minorities, and recruited athletes, among other groups. They also want to appear selective by reporting high test scores, which can lead to a higher ranking and its associated benefits[4]. Under a test-required policy, these inputs may be in conflict. For instance, admitting a legacy student with a low SAT score decreases both the academic preparedness of the student body and the college's perceived selectivity.

By contrast, a test-optional policy may allow a college to achieve its compositional goals while simultaneously appearing more selective, since colleges are only required to report the scores of submitting enrollees and low-scoring applicants may be induced to withhold their scores. However, this is not without sacrifice: while low-scoring applicants who would have been accepted under either regime now choose to withhold their scores, new applicants with very low test scores but competitive non-test observables are incentivized to apply. Unable to distinguish the two, colleges may end up enrolling applicants with very low test scores and high non-test observables at the expense of applicants with intermediate values of both who would have been admitted under a test-required policy.[5] This results in a separation of admitted applicants along the test score dimension; while the college appears outwardly more selective, the average admitted student has a lower test score than before. Kelly (2022) finds suggestive evidence of this: at one selective test-optional college, the average admitted non-submitter had an SAT score at the 67$^{\text{th}}$ percentile, compared to the 90$^{\text{th}}$ percentile for the average admitted submitter. A college may find this outcome favorable if their objective function places sufficient weight on their perceived selectivity and the composition of their student-body.

[4]Reported SAT/ACT scores are an explicit input in the widely consulted *US News & World Report* college ranking. Luca and Smith (2013) show that an as-if random one spot increase in this ranking increases application volume by 1 percent. Highly ranked institutions may also find it easier to solicit donations.

[5]Dessein et al. (2025) develop a formal model of test-optional admissions with similar underpinnings that generates the same theoretical prediction.

This predicted replacement of intermediate scorers with low scorers, combined with the literature on ability sorting across majors, motivates investigating whether test-optional policies cause a decline in STEM completions. The ability sorting literature suggests that the difference between a student's SAT score and the institutional average is a strong, positive predictor of whether they choose a STEM or business major (see Turner and Bowen (1999) and Arcidiacono (2004)); Westrick et al. (2023) find that SAT scores are considerably more predictive of STEM grades than non-STEM grades, suggesting that academic mismatch may be partly responsible. Furthermore, in the context of affirmative action, Arcidiacono et al. (2016) suggest that mismatch only becomes an issue when an individual's preparation is substantially below that of their peers. Consequently, if non-submitters tend to have significantly lower scores than both submitters and displaced students, then one might expect them to struggle in STEM curricula and for STEM completions to decline after policy adoption.

3 Peer Institution Network

3.1 Network Creation

Networks consist of nodes representing agents and edges that connect them. Formally, I model the peer institution network as both *undirected* and *unweighted*; undirected means that node i being connected to node j implies the reverse, and unweighted means that all connections are equally strong. The set of all nodes is given by $V = \{1, \cdots, N\}$, and the set of all edges is given by $E \subseteq \{\{i, j\} \mid i, j \in V,\ i \neq j\}$, which are unordered pairs of nodes. $\{i, j\} \in E$ indicates that node i and node j are directly connected in the network. A walk between node i and k is a sequence of edges from i to k. For instance, if $\{i, j\} \in E$ and $\{j, k\} \in E$, then the sequence $\{\{i, j\}, \{j, k\}\}$ is a walk from i to k. If there is no walk between i and k, then i and k belong to different components of the network.

To construct the peer institution network, I let elements of V correspond to all private colleges and universities. Two institutions i and j are connected, i.e. $\{i, j\} \in E$, if each included the other in a list of peer institutions that they submitted to the National Center for Education Statistic's Integrated Postsecondary Education Data System in reporting year 2020[6]. I will commonly refer to such a relationship as a mutual peer relationship. While one could define connections unilaterally – i.e. i and j are connected if *either* considers the other a peer – defining connections based on *mutual* peer relationships prevents aspirational peer choice from affecting the structure of the network. There are some situations where an institution does not submit a custom set of peers, in which case it is excluded from the network. Figure 1 shows the set of mutual peers for an example institution, The George Washington University.

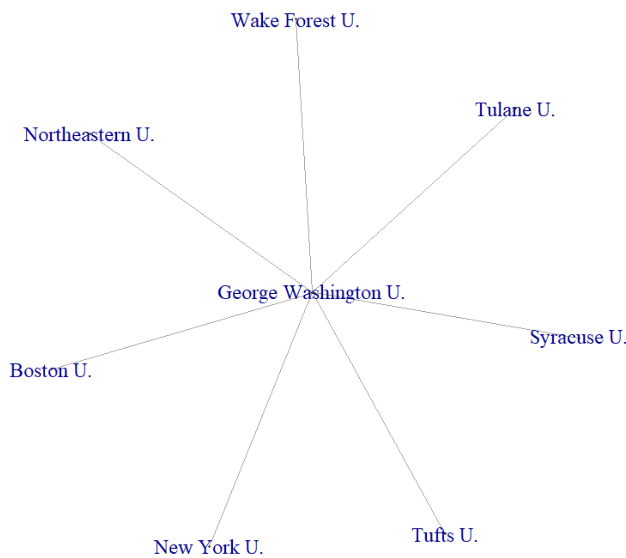
The peer institution network is comprised of one large component, one smaller component, and many singleton or very small components. The purpose of the peer institution network is to identify sets of institutions that can serve as controls for the treated members of their community; communities that are too small will lead to imprecise estimates, so I remove components with 5 or fewer nodes. This leaves the largest and second largest components only, with a combined 617 nodes (institutions) and 1362 edges (mutual peer relationships connecting them).

3.2 Community Detection

Let $\mathcal{C}$ be a partition of V; then, each $c \in \mathcal{C}$ is a collection of nodes from the original network called a community. The community detection problem is to find the "best" partition of nodes, which requires a way of evaluating the quality of any given partition. Newman and Girvan (2004) propose a quality measure they call modularity; for a given partition, modularity captures how many more edges of the graph connect members of the same community than would be expected if the graph were randomly generated. For a community $c \in \mathcal{C}$, let e_c denote the observed number of edges connecting members of c to each other, and let a_c denote how many edges one would expect to connect members of c to each other if the network were generated randomly according to a benchmark network formation model.

[6]Colleges submit peer institution lists in order to receive a "Data Feedback Report" comparing their performance in a variety of metrics to their self-selected peers. See June (2022) for more background on how colleges create their peer institution lists. Peer data were procured from The Chronicle of Higher Education (2020)

Figure 1: Mutual Peers for The George Washington University



Note: Lines reflect that both The George Washington University (GWU) and the connected institution included the other among the set of peer institutions that they submitted to IPEDS in 2020. Figure only includes mutual peer relationships with GWU; many of these institutions are peers with each other as well.

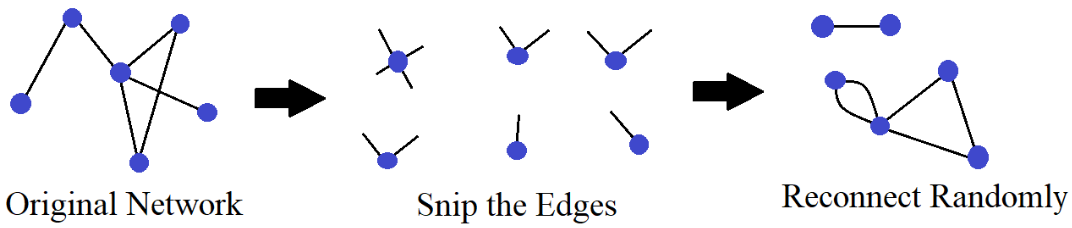
Then, modularity is defined as

$$\mathcal{H}(\mathcal{C}) = \frac{1}{2m} \sum_{c \in \mathcal{C}} (e_c - a_c)\,, \tag{1}$$

where m is the total number of edges in the network and $\frac{1}{2m}$ simply serves as a normalization. As is standard in the community detection literature, I choose the configuration model of Bollobas (1980) to be the benchmark network formation model. Realizations of the configuration model are random "re-wirings" of the original network, generated by "snipping" all the original edges and reconnecting them randomly. Figure 2 depicts this process. This is an attractive benchmark because it generates a random network where each node has the same number of connections as in the original network.

To gain intuition for Equation (1), fix a community c. If $(e_c - a_c)$ is positive, the total number of observed intra-community edges exceeds what would be expected if the network was a realization of the configuration model. Therefore, high values of $\mathcal{H}(\mathcal{C})$ correspond to partitions of nodes into communities that have high intra-community edge density, and consequently lower inter-community edge density.

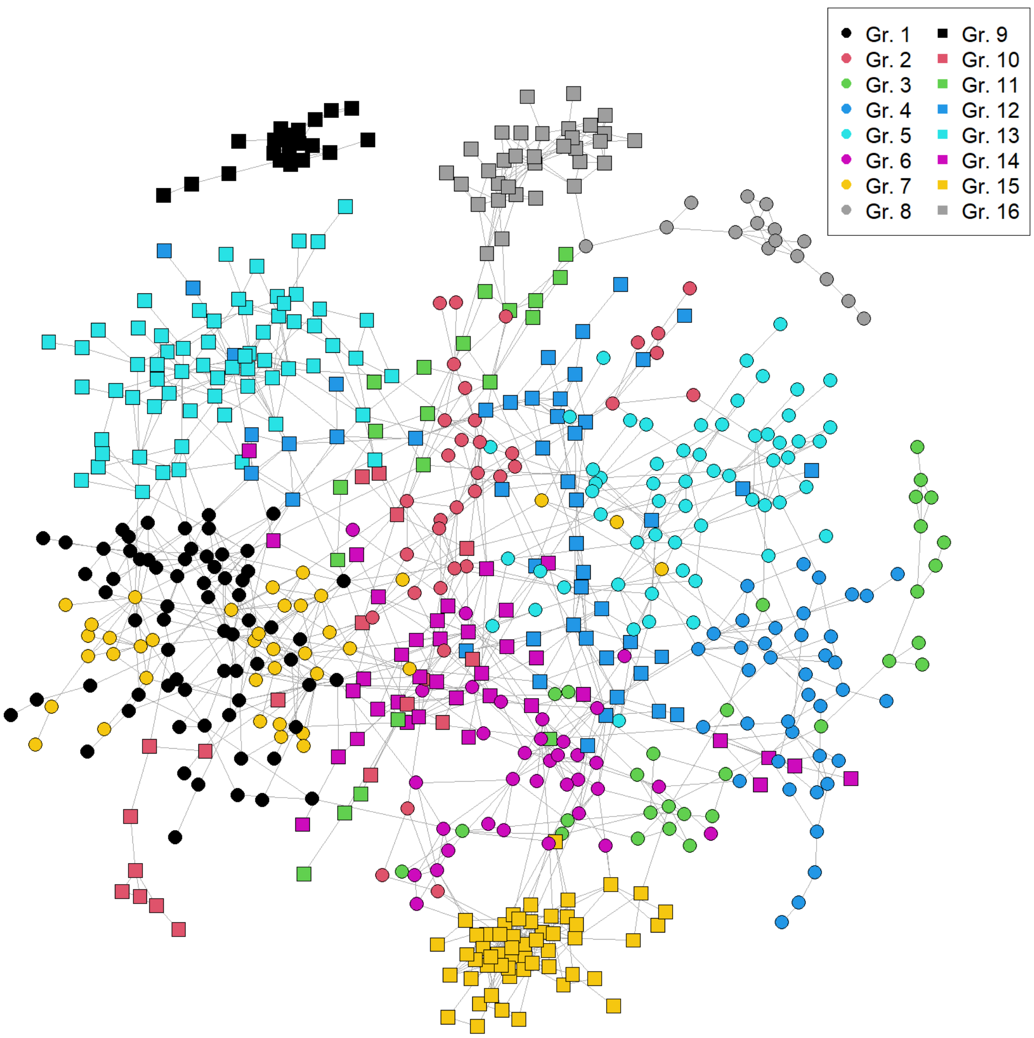
Figure 2: Configuration Procedure



Brandes et al. (2008) show that finding a global maximum of modularity is NP-Hard; therefore, for sufficiently large networks, researchers must rely on heuristic algorithms to find local maxima of modularity. For this paper, I use the Leiden Community Detection algorithm of Traag et al. (2019). It detects 16 communities in the peer institution network. Figure 3 gives a visual depiction of the communities in the peer institution network; nodes sharing the same color and shape combination belong to the same community. A comprehensive list of communities and their members can be found in the Online Appendix.

While there is certainly commonality between the detected communities and, for instance, the US News & World Report college rankings, it is clear that the network structure is capturing something richer than what could be obtained by segmenting the sample by selectivity, size, or institution type. For example, Clark University is considered a National University according to U.S. News & World Report, but is found in community 15 among selective national liberal arts colleges; this placement is capturing the fact that Clark considers its peers to be elite national liberal arts colleges, not elite national universities. Similarly, Harvey Mudd College, a liberal arts college according to US News & World Report, is in community 8, which is comprised mainly of technical colleges and universities like Rochester Institute of Technology and Worcester Polytechnic Institute. This makes sense when you consider that despite its status as a liberal arts college, Harvey Mudd College has a STEM-focused curriculum. Of course, one with sufficiently strong institutional knowledge might well have placed Clark University among liberal arts colleges and Harvey Mudd College among STEM-focused institutions; however, the community detection algorithm is able to make these distinctions without relying on a researcher's input or subjective guidance.

Figure 3: Communities in the Peer Institution Network



Note: Gr. = Group/Community. Dots and squares (nodes) represent colleges and lines connecting them (edges) indicate that each institution included the other among the set of peer institutions that they submitted to IPEDS in 2020. Nodes sharing the same color and shape combination belong to the same community.

The Leiden algorithm will detect communities in *any* network, regardless of whether there is an underlying community structure; in Appendix A, I apply a formal test to demonstrate that the peer institution network exhibits an incredibly strong community structure.

4 Data

4.1 Policy Dates

Test-optional policy start dates are obtained from Bennett (2022), which in many cases have been verified directly with admissions offices. Some policies are conditional, that is to say that an applicant is given the non-submission option as long as they meet some requirement, like exceeding a high-school GPA threshold. Following Bennett (2022), I include these conditional policies in my analysis based on the assumption that the conditions are typically lenient and therefore are satisfied by most applicants. Bennett (2022) also requires that treated institutions are sufficiently selective; a test-optional policy would not matter for an institution that accepts all students. In addition, public and "special-focus" institutions are excluded; the latter include (but are not limited to) art schools and colleges for religious training, and are institutions for which college entrance exams likely play a more limited role in the selection criteria. There are some instances where a college adopts and then rescinds its test-optional policy in a later year; I remove these schools from my sample. I also exclude colleges that disproportionally serve historically marginalized groups, such as Historically-Black Colleges and Universities and Women-only Colleges, as under-represented or female enrollment will change very little over the sample period, making it unlikely that the parallel trends assumption holds.

During my sample period, a number of institutions began accepting the Common Application (CA) from applicants. I obtain CA adoption dates from Knight and Schiff (2022), who find that CA adoption increases application volume, decreases acceptance rates, increases out-of-state enrollment, and may increase the 25^{th} and 75^{th} percentiles of the distribution of reported SAT scores for enrolled students. Consequently, an institution's decision to accept the CA may confound the effect of adopting a test-optional policy; therefore, I restrict my sample to colleges that began accepting the CA before going test-optional, and drop observations prior to CA adoption. To my knowledge, no other paper investigating test-optional policies addresses CA adoption. All sample restrictions happen downstream from the community detection process, so the communities described in the Online Appendix include some institutions that are excluded from the analytic sample.

4.2 Outcome Variables

Almost all outcome data is from the Integrated Postsecondary Education Data System (National Center for Education Statistics (2022)), though I augment some outcomes with hand-collected Common Data Sets from specific colleges and universities for which data is spotty. To exclude the pandemic, I collect outcome data through 2019 (2022 for STEM completions). I drop units treated between 2017 and 2019 and estimate treatment effects up to four periods post-treatment; this ensures that dynamic effects are not confounded by compositional changes across event time.

The selectivity outcomes that I investigate are the 25^{th} and 75^{th} percentile of SAT scores for enrolled and *reporting* students, which I define as the sum of the 25^{th} and 75^{th} percentile Math and Verbal subsection scores. Because of how scores are reported for each, there is substantially more year-to-year variation in reported SAT scores than reported ACT scores, which is why I only investigate the former. Bennett (2022) does not investigate these outcomes, as many treated institutions stopped reporting test score variables to IPEDS following policy adoption. However, they continued to report exactly the same variables in their Common Data Sets, which are typically maintained by their Institutional Research offices. I augment the available IPEDS data with hand-collected historical Common Data Sets. Though I do not get perfect coverage, I fill in at least some of the missing data for 26 treated institutions; post-treatment SAT data is still completely missing for 12 institutions.

I also investigate the log of the share of all graduates with a STEM major and the log of the share of under-represented minority graduates with a STEM major, which I will typically refer to as the overall and URM-only Log Graduate STEM Share, respectively. The URM designation includes Black, Hispanic, and Native American students/graduates. Two-or-more-race and unknown race students are not considered under-represented when those data are available. To my knowledge, this is the first paper to consider the effect of test-optional policy adoption on this outcome. I construct this variable from data on bachelors degree completions broken down by CIP code. To classify a CIP code as STEM or non-STEM, I rely on the Department of Homeland Security's STEM Code classification[7]. Because the DHS list changes over time, I designate a given CIP code as STEM if it belonged to either the 2016 or 2020 DHS STEM Code list. The Graduate STEM Share is given by the sum of all bachelors degree completions in a STEM designated field in a given year divided by the total number of bachelors degree completions in the same year. To align policy dates with graduation cohorts, I lead the Graduate STEM Percent variable by four years. While some students may graduate in fewer than four years, the largest effects should not materialize until at least four years following the policy start year.

Another new outcome that I consider is the percent of enrolled students who did not submit a college entrance exam (the non-submitter percent), which would indicate how willing policy adopters are to enroll non-submitting students. While the non-submitter percent is not directly observable, institutions report the percent of enrolled students who submitted SAT and ACT scores to IPEDS and in their Common Data Sets. I consider the sum of the two percents to be an outcome, hereafter referred to as the Test Submission Percent, with post-policy declines reflecting an increase in the non-submitter percent. Because many students submit scores from both exams, this variable is frequently above 100, especially for not-yet- and never-treated institutions. While imperfect, it nevertheless serves as a reasonable proxy for measuring the extent to which policy adopters are enrolling non-submitting students.

I also collect the following outcomes related to student-body diversity: logged first-time full-time (FTFT) URM enrollment, logged FTFT enrollment of self-identified women, and logged enrollment of FTFT students who are federal-grant recipients. I rely on the enrollment of federal-grant recipients as a proxy for low-income students, as IPEDS does not provide enrollment of Pell-grant recipients until after the start of my sample period[8]. While I do not include the impact of test-optional policies on these outcomes in the main text of this paper,

[7]See 2020 DHS STEM Code List, 2016 DHS STEM Code List

[8]The correlation between FTFT federal-grant recipient and FTFT pell-grant recipient enrollment during the period in which they overlap is greater than 0.9.

they are an interesting dimension across which the network-detected communities differ; I explore this in the next subsection.

4.3 Summary Statistics

In this section I highlight heterogeneity across the communities detected in Section 3; however, to avoid noisy estimates of community-time fixed effects, I only consider communities with 10 or more never-treated members. Table 1 describes these 8 "policy relevant" communities by the number of treated and never-treated units in each and the median treatment year for treated units. Communities differ in terms of treatment uptake. Only 3 of the 23 institutions in community 3 (Less Selective National Universities) adopt a test-optional policy between 2006 and 2016, whereas 16 of the 24 institutions in community 6 (Liberal Arts Colleges - Midwest) do. There is also cross-community variation in treatment timing; Less Selective National Universities and Elite National Universities are the latest adopters, while Colleges and Universities - Midwest are the earliest. In total, I observe 66 treated institutions across the 8 communities. Figure 4 gives a visual depiction of the 8 communities and where they are positioned in the peer institution network.

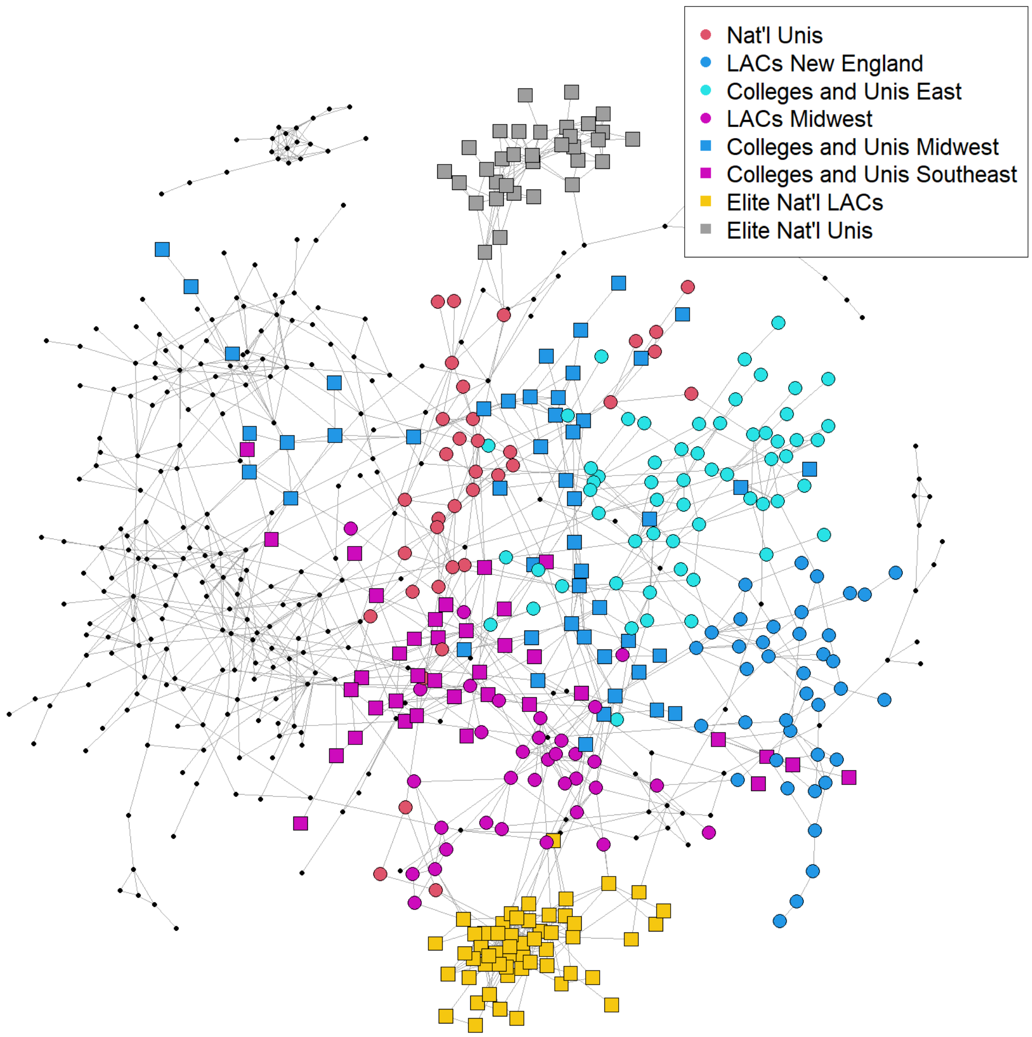
Table 1: Policy Relevant Communities

#	Community Name	Treated	Never-Treated	Median Treatment Year
2	Less Selective National Universities	3	19	2015
4	Liberal Arts Colleges - New England	12	8	2011
5	Colleges and Universities - East	11	20	2011
6	Liberal Arts Colleges - Midwest	16	8	2011
12	Colleges and Universities - Midwest	4	12	2007
14	Colleges and Universities - Southeast	5	22	2009
15	Elite National Liberal Arts Colleges	11	24	2010
16	Elite National Universities	4	26	2013.50

Note: Communities are detected from the Peer Institution Network. Only communities with 2 or more treated institutions are included. An institution is treated if it adopted a test-optional policy between 2006 and 2016, and Never-Treated if it did not adopt a test-optional policy by 2020. Institutions that adopted a policy between 2017 and 2019 are excluded.

Table 2 reports the average reported 25^{th} percentile SAT score, the admit rate, the number of first-time full-time (FTFT) applications, and the number of FTFT enrollees broken down by community and treatment status. The latter two variables are from IPEDS. Summary statistics are calculated using data from 2001 to 2005 only, and therefore excludes treated observations. There is substantial heterogeneity across community. Elite National

Figure 4: Policy Relevant Communities in the Peer Institution Network



Note: Nat'l = National. LAC = Liberal Arts College. Unis = Universities. Dots and squares (nodes) represent colleges and lines connecting them (edges) indicate that each institution included the other among the set of peer institutions that they submitted to IPEDS in 2020. Nodes sharing the same color and shape combination belong to the same community. Highlighted nodes belong to communities with more than one treated institution; communities with one or fewer treated institutions are not used to evaluate test-optional policies.

Table 2: Selectivity and Enrollment by Community and Treatment Status

Community	SAT 25$^{\text{th}}$ %ile		Admit Rate		FTFT Applications		FTFT Enrollment	
	T	NT	T	NT	T	NT	T	NT
Less Selective National Universities	1046	1078.74	75.73	66.14	7185.20	6423.45	1500.54	1208.55
	(63.01)	(56.71)	(8.24)	(16.99)	(3332.02)	(3269.50)	(596.26)	(553.05)
Liberal Arts Colleges - New England	942.70	907.40	71.44	74.95	2774.61	2335.86	483.33	549.11
	(105.83)	(53.04)	(10.71)	(15.74)	(1438.36)	(1977.65)	(214.24)	(380.02)
Colleges and Universities - East	1044.93	932.89	64.79	74.76	5799.75	2358.79	865.87	534.17
	(74.07)	(76.69)	(11.03)	(12.23)	(3249.45)	(1806.01)	(368.72)	(366.21)
Liberal Arts Colleges - Midwest	1049.97	1031.22	69.68	76.66	2022.83	1957.90	387.27	465.77
	(68.13)	(62.45)	(11.88)	(8.09)	(739.09)	(768.48)	(76.02)	(169.16)
Colleges and Universities - Midwest	1091	931.84	69.43	76.74	2084.12	1225.61	444.07	331.36
	(40.28)	(118.28)	(12.42)	(12.97)	(640.36)	(661.92)	(157.57)	(159.38)
Colleges and Universities - Southeast	1016.20	1007	81.31	78.29	2366.88	2341.10	574.10	566.85
	(42.59)	(62.47)	(4.33)	(9.88)	(500.48)	(1329.48)	(113.97)	(239.31)
Elite National Liberal Arts Colleges	1157.15	1251.21	51.60	41.04	4135.05	4152.95	589.33	520.49
	(73.07)	(59.27)	(14.44)	(15.94)	(1291.80)	(1630.74)	(105.25)	(160.30)
Elite National Universities	1218.35	1281.45	44.59	32.44	10165.75	15984.31	1316.87	1839.05
	(34.19)	(86.61)	(3.93)	(16.37)	(4965.57)	(6862.15)	(640.51)	(933.01)
Full Sample	1050.78	1080.60	66.35	63.39	3841.64	4987.92	639.17	771.48
	(107.59)	(164.42)	(14.45)	(22.77)	(3019.66)	(5934.47)	(414.68)	(694.65)

Note: Summary statistics are calculated using data from 2001 to 2005. Standard deviations are in parentheses. FTFT = First-time full-time. T denotes that a column's statistics are conditional on a unit being eventually treated, and NT denotes that a column's statistics are conditional on never being treated.

Liberal Arts Colleges (community 15) and Elite National Universities (community 16) report the highest SAT scores, while institutions in community 4 (Liberal Arts Colleges - New England) report the lowest SAT scores, with reported 25$^{\text{th}}$ percentile SAT scores below the population average. There is some heterogeneity across treatment status and within community; treated Elite National Liberal Arts Colleges and Elite National Universities report 25$^{\text{th}}$ percentile SAT scores that are lower than their never-treated counterparts; however, the heterogeneity goes the other direction in other communities. Differences in Admit Rates are qualitatively similar to the differences in standardized test scores. Communities differ substantially in terms of size, with Elite National Universities and Less Selective National Universities (community 2) enrolling the most FTFT students, while community 12 (Colleges and Universities - Midwest) is comprised of smaller schools with average FTFT enrollment around 400.

Table 3: Student Demographics by Community and Treatment Status

Community	FTFT % URM		FTFT % Women		FTFT Fed. Grant %		Overall Graduate STEM %		URM-only Graduate STEM %	
	T	NT	T	NT	T	NT	T	NT	T	NT
Less Selective National Universities	12.79	12.48	57.66	55.69	22.08	17.18	12.92	16.18	10.94	13.48
	(7.61)	(5.82)	(3.26)	(6.95)	(4.96)	(5.55)	(5.12)	(7.74)	(4.80)	(7.62)
Liberal Arts Colleges - New England	6.05	9.88	56.41	53.68	20.10	28.92	9.22	10.80	10.00	11.28
	(5.15)	(8.72)	(12.26)	(16.17)	(8.91)	(12.56)	(5.22)	(17.82)	(11.70)	(20.54)
Colleges and Universities - East	7.94	12.85	57.54	61.45	16.48	31.86	10.70	8.84	9.24	7.06
	(3.94)	(8.93)	(11.78)	(12.88)	(7.68)	(16.64)	(3.88)	(4.54)	(4.86)	(5.64)
Liberal Arts Colleges - Midwest	6	6.62	57.25	53.23	18.31	20.19	18.00	18.34	13.58	12.22
	(3.33)	(3.49)	(5.69)	(7.01)	(7.51)	(8.35)	(6.82)	(3.62)	(11.02)	(9.52)
Colleges and Universities - Midwest	6.91	8.26	56.64	55.69	19.36	35.68	19.38	12.78	13.14	11.80
	(3.48)	(8.69)	(3.28)	(10.12)	(4.62)	(22.44)	(3.86)	(6.48)	(8.82)	(15.94)
Colleges and Universities - Southeast	6.80	8.12	60.17	57.41	23.73	24.73	14.12	14.10	8.54	10.08
	(2.30)	(5.17)	(8.15)	(7.07)	(15.79)	(12.77)	(5.30)	(7.78)	(5.52)	(8.44)
Elite National Liberal Arts Colleges	7.90	9.92	54.69	53.35	13.17	11.88	15.84	20.08	11.78	13.74
	(3.03)	(5.24)	(3.73)	(4.02)	(4.32)	(6.30)	(2.86)	(5.90)	(7.02)	(7.40)
Elite National Universities	8.36	13.81	52.68	49.97	13.61	15.23	17.88	28.08	13.18	23.20
	(1.87)	(4.99)	(4.85)	(6.49)	(6.97)	(6.78)	(8.58)	(20.78)	(7.40)	(20.52)
Full Sample	7.20	10.49	56.67	55.33	18.03	24.06	14.30	16.28	11.36	13.10
	(4.34)	(7.27)	(8.53)	(9.77)	(8.51)	(16.50)	(6.46)	(12.40)	(9.04)	(14.06)

Note: Summary statistics are calculated using data from 2001 to 2005. Standard deviations are in parentheses. FTFT = First-time full-time. URM = under-represented minority. Graduate STEM % = Percent of graduates majoring in a STEM field. T denotes that a column's statistics are conditional on a unit being eventually treated, and NT denotes that a column's statistics are conditional on never being treated.

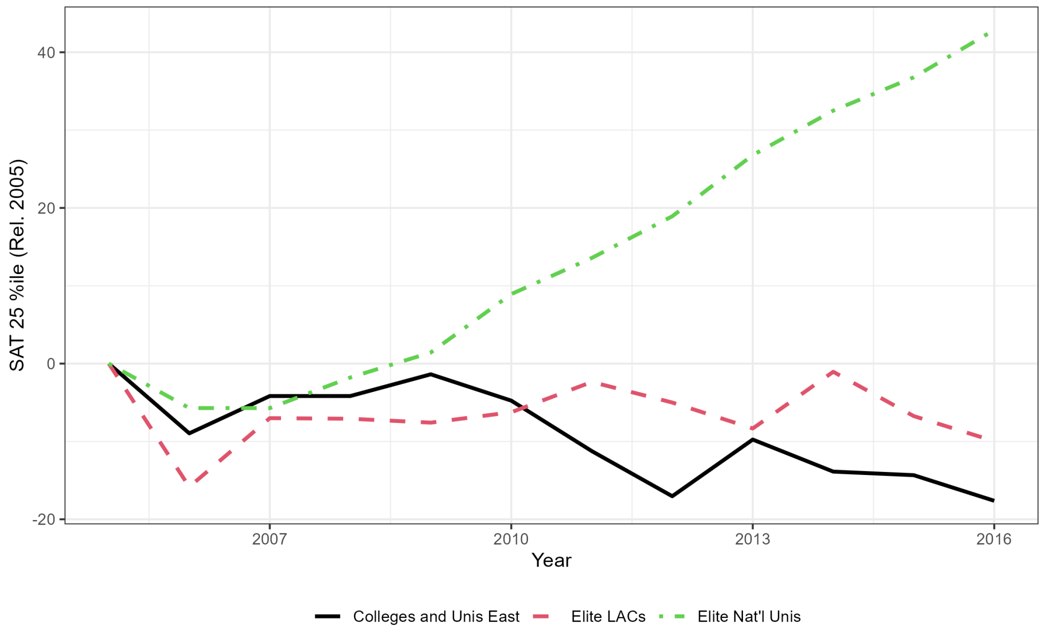
Table 3 reports student demographics by community and treatment status. Variables include the percent of FTFT students who are under-represented minorities, women, and federal grant recipients, and the overall and under-represented minority-only Graduate STEM Share (expressed as a percent). As with Table 2, the data are from 2001 to 2005. Community 2 (Less Selective National Universities) and community 16 (Elite National Universities) are the most racially diverse. On average, treated institutions are less racially and socioeconomically diverse than their never-treated counterparts, which may suggest that improving student-body diversity is the key motivation for adopting a test-optional policy. There is less variation across community in the percent of FTFT students who are women, who seem to outnumber men in every community and are more overrepresented at treated institutions. There is some variation in the percent of students receiving federal grants, which is my proxy for Pell-grant receiving students. Elite National Liberal Arts Colleges and Elite National Universities enroll relatively fewer federal grant recipients than the other communities. The overall Graduate STEM Share varies a little across communities, but hovers between 10 and 20 percent for most. Almost without exception, the mean Graduate STEM Share is lower for under-represented minorities than for the overall student-body.

4.4 Community-Specific Trends

If treated and untreated community members share the same counterfactual trends, population-level treatment effects can be estimated based on a version of the parallel counterfactual trends assumption that is conditional on community membership; however, this assumption is unnecessary if trends do not differ by community. Figure 5 plots trends in the reported 25^{th} percentile SAT scores at three communities: community 5 (Colleges and Universities - East), community 15 (Elite Liberal Arts Colleges), and community 16 (Elite National Universities). Only never-treated institutions are used to calculate the trend. It is clear that the elite national universities are on a substantially different trend than the other two communities. Elite national universities report 25^{th} percentile SAT scores more than 40 points higher in 2016 than in 2005, while the other two communities see declines in the 25^{th} percentile of reported SAT scores.

When estimating population-level treatment effects, these differences can lead to biased

Figure 5: Trends in Reported 25th Percentile SAT Scores for Selected Communities



Note: Unis = Universities. LAC = Liberal Arts College. Nat'l = National. Time-series represents the trend in the average 25th percentile SAT score of enrolled students at never-treated members of each selected community. Trends are relative to the average 25th percentile SAT score in each community in 2005.

estimates. For instance, assuming that treated members of Colleges and Universities - East would have evolved in the same way as their never-treated counterparts, including Elite National Universities as controls would lead to a severe downward bias in the estimated effects on reported SAT scores, as the underlying trends between the two groups are diverging. A counterfactual trends assumption that is conditional on community avoids this problem. Of course, many papers, including Bennett (2022), use techniques like Rosenbaum and Rubin (1985)'s propensity score matching to balance the control and treated groups; however, if treatment timing differs by community, then estimates can still be biased. For instance, if treated Elite National Universities become treated later than other treated institutions (Table 1 indicates this to be the case), they will be over-represented in the control sample for units that are treated earlier, introducing a similar downward bias. Again, if a parallel trends assumption holds conditional on community membership, community-specific trends eliminate this bias.

5 Empirical Strategy

When treatment timing is staggered, heterogeneous and/or delayed effects can introduce bias and complicate interpretation when treatment effects are estimated by two-way fixed-effect regression. These issues are discussed at length in many papers, including Borusyak et al. (2024), Gardner et al. (2024), Goodman-Bacon (2021), Callaway and Sant'Anna (2021), and Sun and Abraham (2021). To ensure estimation of easily interpretable impacts of test-optional policy adoption, I employ the two-stage difference-in-differences estimator of Gardner et al. (2024)[9].

Let y_{ct} be the outcome of interest for college c observed in period t. In the first stage, all treated observations (observations that fall after a unit has been treated) are removed, and institution and *community-time* fixed effects are estimated, i.e. the following model is estimated with pre-treatment data only:

$$y_{ct} = \gamma_c + \sum_{g \in \mathcal{G}} \gamma_{gt} \mathbf{1}\{c \in g\} + \nu_{ct}. \tag{2}$$

In Equation (2), g represents a community in the set of all communities $\mathcal{G}$, γ_c is an institution-specific component, and γ_{gt} is a community-time specific component. I assume that Equation (2) describes the data generating process for y_{ct} in the absence of treatment; in the potential outcomes framework of Rubin (1974), $\gamma_c + \gamma_{gt}$ is the never-treated potential outcome for college c in period t.

Let $\hat{\gamma}_c$ and $\hat{\gamma}_{gt}$ be estimates of γ_c and γ_{gt} obtained from the first stage. Then, $d_{ct} = y_{ct} - \hat{\gamma}_c - \hat{\gamma}_{gt}$ is the deviation of y_{ct} from where it would have been in expectation had it followed the process described by Equation (2). Also, let T_c represent the period that college c initially becomes treated. In the second-stage, I calculate d_{ct} for all treated and never-treated observations and regress it on leads and lags from policy adoption to estimate

[9] This procedure produces identical treatment effect estimates to the "imputation" estimator of Borusyak et al. (2024) when treated units are given equal weight; however, the two approaches differ in their asymptotic theories and corresponding variance estimators.

dynamic treatment effects, estimating

$$d_{ct} = \sum_{k=-a}^{b} \tau_k \mathbf{1}\{t - T_c = k\} + \tau_{a^-} \mathbf{1}\{t - T_c < -a\} + \varepsilon_{ct}, \tag{3}$$

where $\mathbf{1}\{t - T_c = k\}$ is an indicator function that takes value 1 if a unit was first treated exactly k periods ago and $\mathbf{1}\{t-T_c < -a\}$ takes value 1 if college c will be treated in more than a periods; therefore, τ_k for $k \in \{-a, \cdots, b\}$ represent period-level dynamic treatment effects and τ_{a^-} represents the treatment effect more than a periods prior to treatment. Estimates of τ_k reflect the average deviation from the estimated never-treated potential outcome for *treated* units in the k^{th} treatment period; therefore, γ_k are estimates of the Average Treatment Effect on the Treated. If there is no treatment anticipation and treated units follow community-specific trends prior to treatment, τ_k should be zero when k is negative[10].

I investigate whether there is treatment effect heterogeneity across communities by estimating the following equation:

$$d_{ct} = \sum_{g \in \mathcal{G}} \tau_g \mathbf{1}\{t - T_c \geq 0 \wedge c \in g\} + \varepsilon_{ct}. \tag{4}$$

Here, $\mathbf{1}\{t - T_c \geq 0 \wedge c \in g\}$ is an indicator that takes value 1 if college c has been treated and belongs to community g. In practice, I estimate the τ_g by regressing d_{ct} on interaction terms between a post-treatment dummy and dummy variables for community membership. Estimates of τ_g reflect the average deviation from the estimated never-treated potential outcome for treated units in community g across all post-treatment periods and therefore reflect Average Treatment Effects on the Treated in community g.

To formally test whether there is treatment effect heterogeneity across communities, I perform an F-test to compare the unrestricted model where effects are community specific against a restricted model where effects are assumed to be equal across communities; this comparison implies a null hypothesis that effects are equal across communities. I calculate the relevant F-statistic and report the corresponding p-value. For simplicity, I estimate

[10]One feature of the Gardner et al. (2024) estimator is that $\hat{\tau}_{-1} \neq 0$, whereas most event studies make the normalization $\hat{\tau}_{-1} = 0$. To interpret estimates in the "usual" way, one can simply compare post-treatment effect sizes to the last pre-treatment estimate. I do this several times in Section 6 to facilitate interpretation.

the restricted and unrestricted models by OLS; however, all reported treatment effects are obtained from the two-stage difference-in-differences procedure. I also selectively document pairwise heterogeneity across communities with a standard Student's t-test using the reported effect sizes and standard errors.

Because d_{ct} is a generated regressor, the standard errors must be corrected as in Gardner et al. (2024) . I use the `R` package `did2s` from Butts and Gardner (2021) to perform the two-stage procedure and calculate standard errors, which are clustered at the institution level. I estimate dynamic treatment effects for the first four treatment periods only; more distant effects may reflect unobserved policy changes and not the impact of the policy of interest.

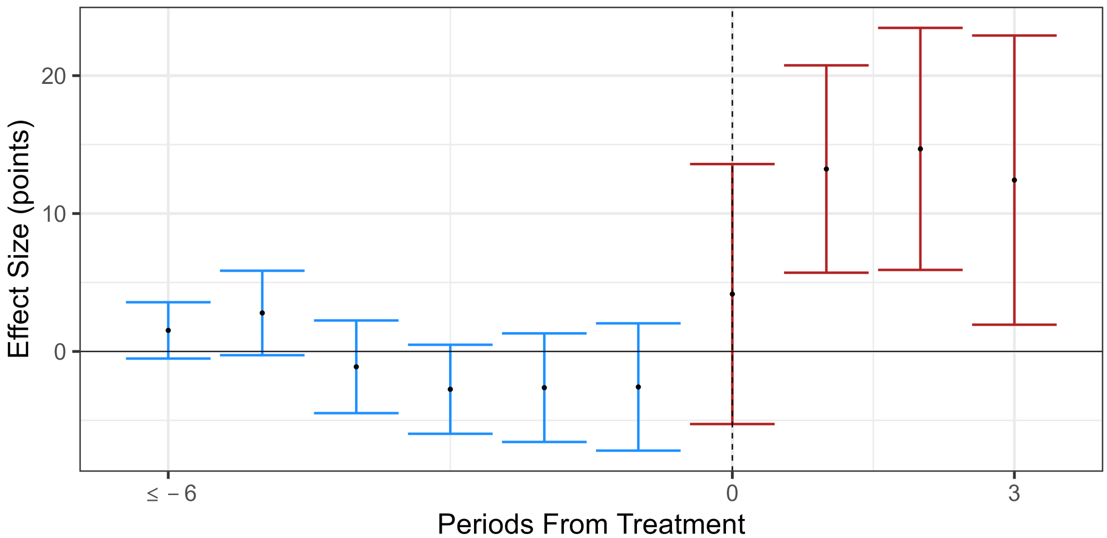
6 Results

6.1 Dynamic Effects (Event-Studies)

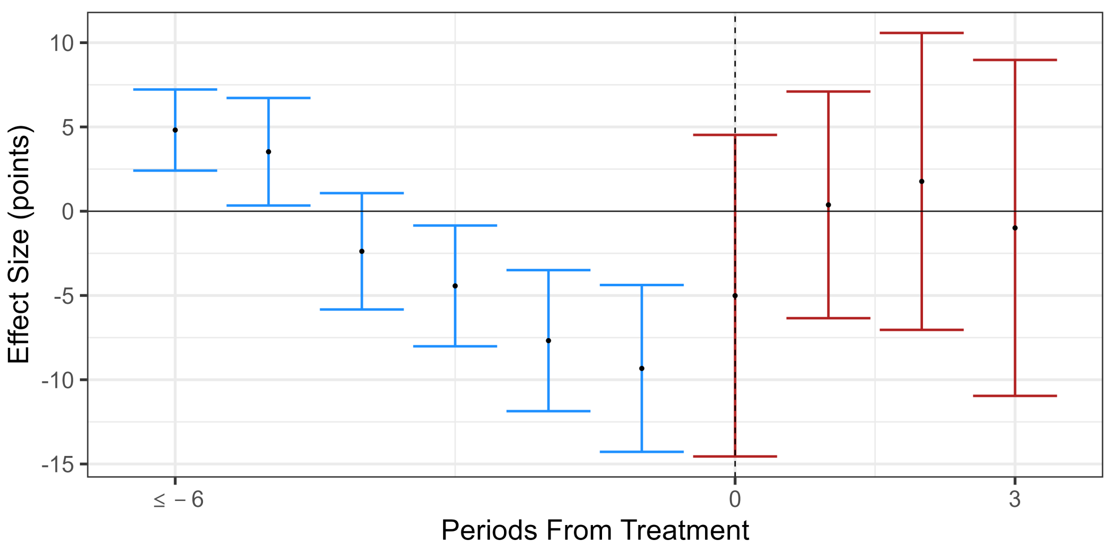
I start by highlighting the advantage of estimating community-specific time fixed effects ($\hat{\gamma}_{gt}$) rather than population-level time fixed effects ($\hat{\gamma}_t$) in Equation (2). Consider the 75$^{\text{th}}$ percentile of reported SAT scores as the outcome of interest. Figure 6 panel (a) presents estimated dynamic treatment effects when time trends are community specific, and Figure 6 panel (b) presents estimated dynamic treatment effects when time trends are estimated at the population-level. In panel (a), the pre-treatment estimates are small in magnitude and stable, indicating that, conditional on community, treated and never-treated institutions evolve similarly prior to treatment. In contrast, there is a severe negative pre-trend in panel (b), indicating that, prior to treatment, the 75$^{\text{th}}$ percentile of reported SAT scores is growing more quickly at control institutions than at treated institutions. Post-treatment estimates in panel (a) are positive and statistically significant, capping out at just under 15 points, but are negative on average in panel (b). If estimates were relative to the final pre-treatment period, they would still be more than 6 points higher in panel (a) than in panel (b), suggesting that the disparate trends is introducing the negative bias discussed in Section 4.4. Therefore, unless otherwise specified, all treatment effect estimates in the main body of the paper

Figure 6: Dynamic Estimates of Policy Adoption on 75$^{\text{th}}$ %ile SAT Score

(a) Model With Community-Specific Time Fixed Effects



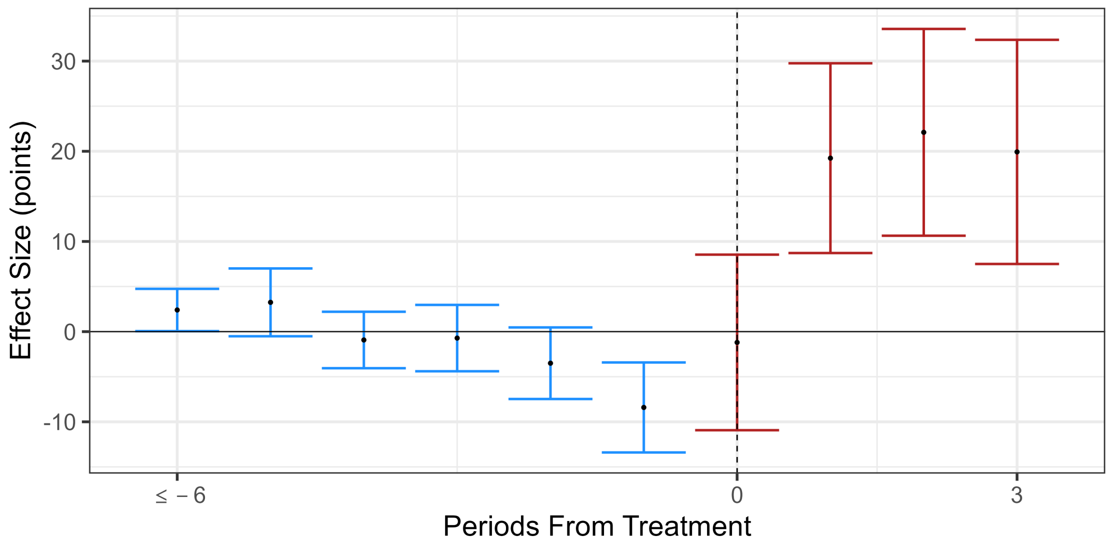
(b) Model With Population-Level Time Fixed Effects



Note: Coefficients represent deviations from the counterfactual potential outcome. Standard errors are from Gardner et al. (2024) and clustered at the institution level. Error bars represent 90 percent confidence intervals. Coefficients are in terms of SAT points. Point estimates for panel (a) and (b) can be found in Appendix B.

will come from models featuring community-specific time fixed effects. Treatment effects estimated from models with population-level time fixed effects can be found in Appendix B. In Appendix C, I consider an alternative grouping of colleges based on their Carnegie Classification (hereafter CC) and estimate dynamic effects with a model where community-trends are based on those groups; network-detected communities still perform better from a pre-trends perspective and provide different results for some outcome variables.

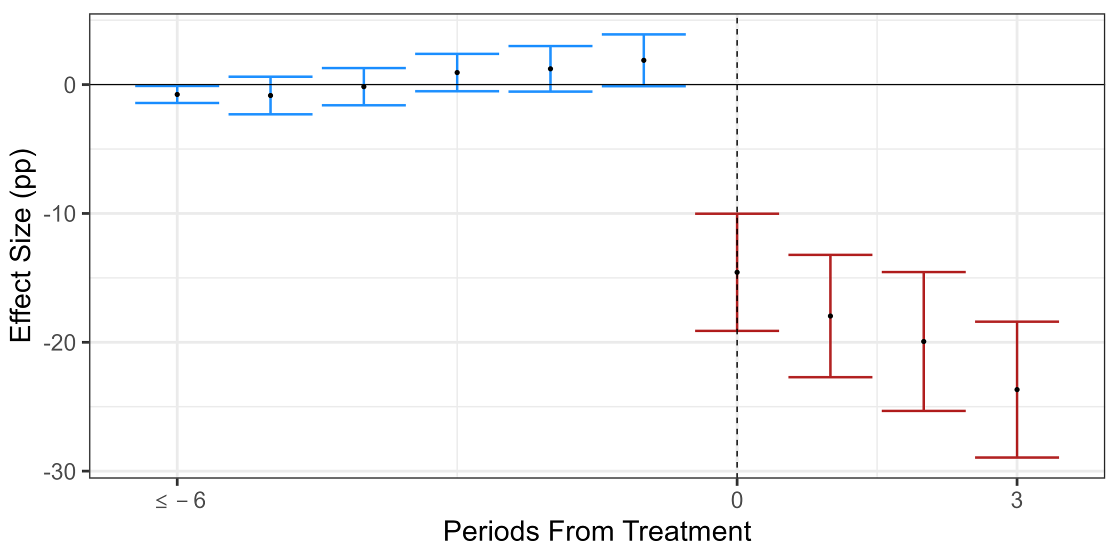
Figure 7: Dynamic Effects of Policy Adoption on 25$^{\text{th}}$ %ile SAT Score



Note: Estimates from model with community-specific trends. Coefficients represent deviations from the counterfactual potential outcome. Standard errors are from Gardner et al. (2024) and clustered at the institution level. Error bars represent 90 percent confidence intervals. Point estimates found in Appendix B.

Figure 7 presents dynamic treatment effect estimates for the 25$^{\text{th}}$ percentile of reported SAT scores. Reported SAT scores for enrolled and submitting students increase substantially following policy adoption. Comparing Figure 7 to Figure 6 panel (a), effects are larger for the 25$^{\text{th}}$ percentile of reported scores than the 75$^{\text{th}}$ percentile, with effect sizes of more than 20 points (estimated effects are 8 points larger if compared to the last pre-treatment estimate). While the pre-trend on the 25$^{\text{th}}$ percentile of reported scores is less substantial when using within-community comparisons only (see Appendix Table B2 for the alternative), it still warrants discussion. As I discuss in the next section, the pre-trend is driven mainly by the communities of Less Selective and Elite National Universities, who both exhibit strong negative pre-trends in the reported test score variables. Excluding these communities compresses the pre-treatment estimates towards zero and increases each of the post-treatment estimates by over 5 points. This suggests that – in the context of test-optional policy adoption – within community comparisons might lead to consistent estimation of treatment effects for some community/outcome combinations and not others. In any case, the direction of the pre-trend would serve to downwardly bias my estimates, so the estimated effects are, if anything, an underestimate of the true effect of policy adoption on reported scores. Strong negative pre-trends in the SAT score variables indicates that reported scores were increasing more quickly (or decreasing less quickly) at untreated members than at treated members of the same community, which may explain the motivation to adopt a test-optional policy.

Figure 8: Effect of Policy Adoption on Test Submission Percent

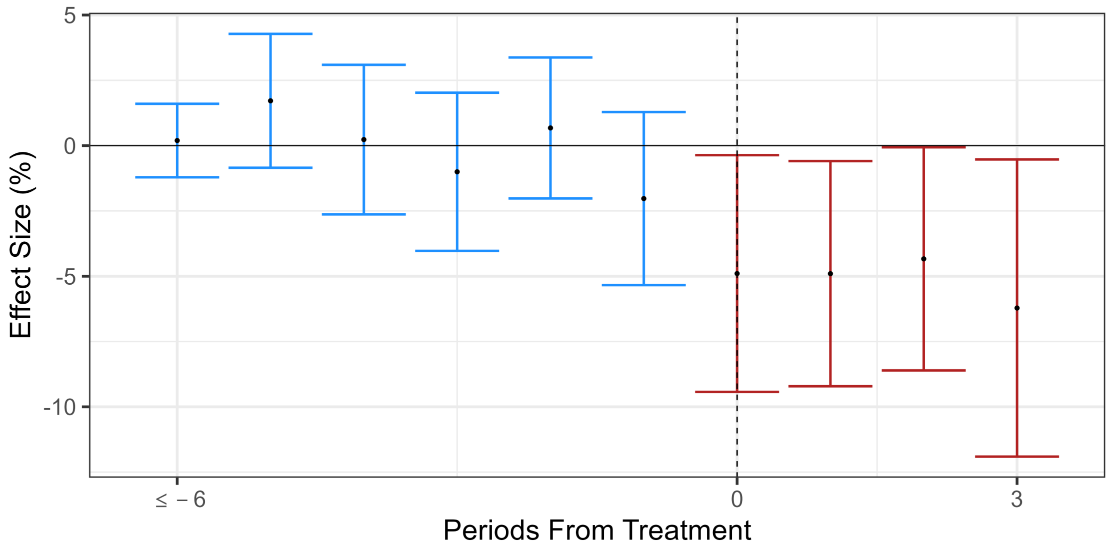


Note: Estimates from model with community-specific trends. Coefficients represent deviations from the counterfactual potential outcome. Standard errors are from Gardner et al. (2024) and clustered at the institution level. Error bars represent 90 percent confidence intervals. Point estimates found in Appendix B.

Figure 8 presents dynamic treatment effects of policy adoption on the Test Submission Percent. Effect sizes start at around -15 percentage points and decrease to around -25 percentage points by the fourth treatment period. These estimates suggest that test-optional schools do enroll a substantial fraction of non-submitters; assuming that the percent of enrolled students submitting both SAT and ACT scores is unaffected by policy adoption, the entire decline in the Test Submission Percent represents non-submitters. All dynamic estimates for the test score variables can be found in Appendix B.

Figure 9 presents dynamic effects of policy adoption on the overall Log Graduate STEM Share. All four post-treatment estimates are negative and statistically significant at the

Figure 9: Effect of Policy Adoption on the Overall Log Graduate STEM Share



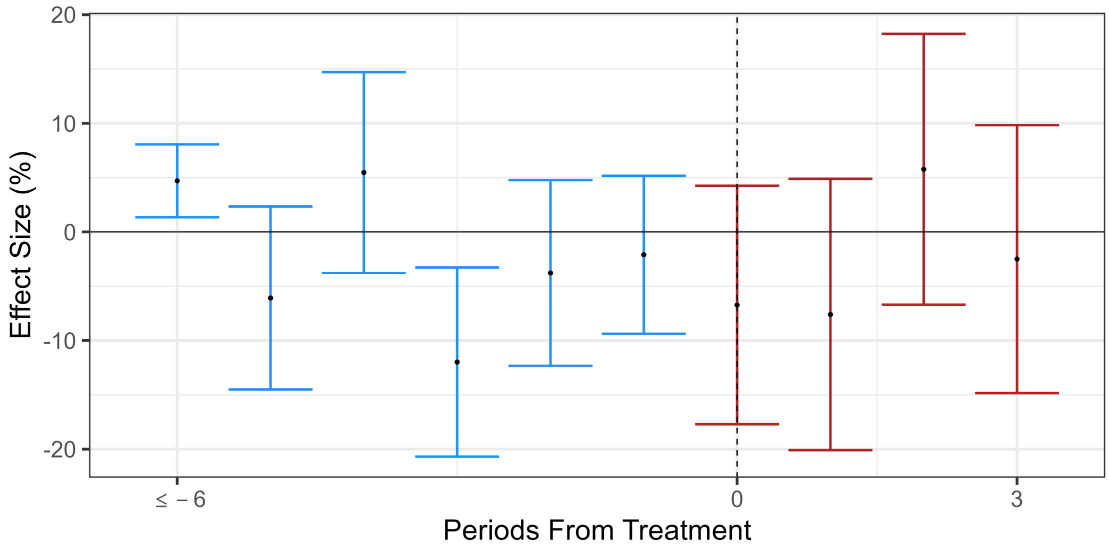
Note: Estimates from model with community-specific trends. Coefficients represent deviations from the counterfactual potential outcome. Standard errors are from Gardner et al. (2024) and clustered at the institution level. Error bars represent 90 percent confidence intervals. Point estimates found in Appendix B.

10% level. In the fourth post-treatment period, the Graduate STEM Share is more than 6 percent lower than in the never-treated counterfactual. Regressing d_{ct} on a single post-treatment indicator yields a point estimate of -0.05, significant at the 5% level. I take this as compelling evidence that policy adoption leads to a decline in the share of graduates with a STEM major. Figure 10 presents analogous effects for the under-represented minority-only Log Graduate STEM Share. Post-treatment estimates are noisy and indistinguishable from zero. Point estimates can be found in Appendix B.

6.2 Community-Level Estimates

Table 4 presents community-level static treatment effect estimates for the two reported SAT score variables and the Test Submission Percent. Effects on the reported SAT score variables are positive for all communities except Less Selective and Elite National Universities. In Appendix E, I show that these negative estimates are driven by a strong negative pre-trend, suggesting that treated members of these communities may have been motivated to adopt a test-optional policy by declining selectivity relative to their peers. Thus, the assumption of parallel counterfactual trends is unlikely to be satisfied in those communities for these

Figure 10: Effect of Policy Adoption on the URM-only Log Graduate STEM Share



Note: Estimates from model with community-specific trends. URM = Under-represented minority. Coefficients represent deviations from the counterfactual potential outcome. Standard errors are from Gardner et al. (2024) and clustered at the institution level. Error bars represent 90 percent confidence intervals. Point estimates found in Appendix B.

particular outcomes. Other communities see more stable pre-trends.

Colleges and Universities - East and Colleges and Universities - Midwest see the largest increases in the 25$^{\text{th}}$ percentile of reported SAT scores, with an effect size around 31 points, though the latter estimate is noisy. Across community, effects are larger for the 25$^{\text{th}}$ percentile of reported scores, with little exception. Liberal Arts Colleges - New England see the most significant effects, with both estimates positive and significant at the 5% level. However, after excluding the Less Selective and Elite National Universities, I can not reject that effect sizes are the same across community. In general, communities see a statistically significant decline in the Test Submission Percent, but the declines are smallest for Less Selective National Universities, Elite National Universities, and Liberal Arts Colleges - Midwest; the other five communities see declines over 20 percentage points. The decline is largest at Liberal Arts Colleges - New England, who see a 32 percent decline in the Test Submission Percent, with a 95% confidence interval of [-39.24, -25.12]. Accordingly, I can reject at all three conventional levels that declines in the Test Submission Percent are equal across communities.

Table 5 reports community-level static treatment effects for the Log Graduate STEM Share overall and by under-represented minority status. There is considerable heterogeneity across communities. For the overall Graduate STEM Share, Colleges and Universities -

Table 4: Effects on Test Score Variables by Community

Outcome:	SAT 25th %ile	SAT 75th %ile	Test Submission Percent
Model:	(1)	(2)	(3)
Community			
Less Selective Nat'l Unis	-18.78**	-1.642	3.015
	(9.236)	(6.548)	(4.983)
LACs - New England	25.23***	25.45**	-32.18***
	(9.644)	(11.22)	(3.602)
Colleges and Unis - East	30.99**	22.69	-26.15***
	(14.03)	(15.13)	(4.874)
LACs - Midwest	13.37	8.925	-10.89*
	(10.22)	(9.091)	(5.648)
Colleges and Unis - Midwest	31.17	10.13	-25.61***
	(32.83)	(12.88)	(6.527)
Colleges and Unis - Southeast	21.03	11.17	-23.11***
	(20.48)	(15.10)	(6.450)
Elite Nat'l LACs	13.54	4.376	-22.47***
	(14.48)	(10.09)	(5.412)
Elite Nat'l Unis	-31.05***	-6.418	-7.347**
	(9.534)	(11.42)	(3.668)
Prob. All Effects Equal	<0.01	0.093	<0.01
Fit statistics			
Observations	2,700	2,700	2,518
R^2	0.04556	0.02310	0.27751
Adjusted R^2	0.04308	0.02056	0.27550

Custom standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

East and Elite National Liberal Arts Colleges see statistically significant declines, with point estimates of -0.1571, and -0.0837, and both are statistically significant at the 1% level. The other communities see no significant change in the Log Graduate STEM Share following adoption. Effect sizes are less precisely estimated for the URM-only Graduate STEM Share, where declines are concentrated in the same two communities; following policy adoption, Colleges and Universities - East and Elite National Liberal Arts Colleges see 26 and 23 percent declines, on average, in the proportion of under-represented minorities completing bachelor's degrees in STEM, with the latter effect significant at the 1% level. Across all three outcomes (overall, URM, and non-URM Graduate STEM Share), I can reject at the 5% level that effects are the same across communities.

As with the negative effect on reported SAT scores at Less Selective and Elite National

Table 5: Effects on Graduate STEM Share

Outcome:	Log Graduate STEM Share		
	Overall	URM Only	non-URM Only
	(1)	(2)	(3)
Community			
Less Selective Nat'l Unis	0.0565	0.0191	0.0568
	(0.1363)	(0.0749)	(0.1471)
LACs - New England	-0.0917	0.0901	-0.0979
	(0.0845)	(0.1970)	(0.0884)
Colleges and Unis - East	-0.1571***	-0.2646*	-0.1511***
	(0.0590)	(0.1423)	(0.0541)
LACs - Midwest	0.0019	0.1151	-0.0070
	(0.0395)	(0.0879)	(0.0384)
Colleges and Unis - Midwest	-0.0973	0.0373	-0.1016*
	(0.0602)	(0.1922)	(0.0591)
Colleges and Unis - Southeast	0.0394	0.0430	0.0267
	(0.0350)	(0.1283)	(0.0401)
Elite Nat'l LACs	-0.0837***	-0.2300***	-0.0664**
	(0.0314)	(0.0786)	(0.0314)
Elite Nat'l Unis	0.0539	0.0516	0.0597
	(0.0878)	(0.0892)	(0.0884)
Prob. All Effects Equal	<0.01	0.019	<0.01
Fit statistics			
Observations	2,714	2,714	2,714
R^2	0.02687	0.01659	0.02415
Adjusted R^2	0.02435	0.01405	0.02162

Custom standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Universities, statistically significant static effects could be driven by different underlying trends between treated and control institutions; in Appendix E, I present estimates from a regression of d_{ct} (deviations from counterfactual potential outcomes) on treatment leads plus a post-treatment indicator, allowing me to determine if treated institutions were evolving similarly to control institutions prior to treatment. I find that the significant negative effect of policy adoption on the overall and URM-only Graduate STEM Share at Colleges and Universities on the East Coast and Elite National Liberal Arts Colleges is not explained by different underlying trends between treated and control units; estimates are stable and near zero prior to treatment, dropping sharply thereafter.

7 Discussion and Conclusion

In this paper, I exploit a peer institution network to identify communities of colleges and universities. Communities differ substantially in selectivity, size, and student-body diversity. I use these detected communities to estimate both the population- and community-level impacts of test-optional policies, relying on a parallel counterfactual trends assumption that is conditional on community membership.

Across 66 institutions that adopted a test-optional policy between 2006 and 2016, I find significant increases in the perceived selectivity of adopting institutions, as measured by the reported SAT scores of enrolled students. The 25$^{\text{th}}$ and 75$^{\text{th}}$ percentiles of the reported SAT score distribution are 22 and 15 points higher, respectively, in the third post-treatment period than in the never-treated counterfactual. Increases are largest among Liberal Arts Colleges in New England and Colleges and Universities on the East Coast, where point estimates on the 25$^{\text{th}}$ (75$^{\text{th}}$) percentile are 25 and 31 (25 and 23) points, respectively; however, after excluding two communities for which pre-trends complicate identification, I cannot reject that effect sizes are equal across communities. Interestingly, the two communities who see the largest increase in reported scores also see the largest decrease in the Test Submission Percent, a proxy for the proportion of non-submitting students. Taken together, these results are consistent with the prediction described in Section 2 and by Dessein et al. (2025); test-optional policies lead to a left censoring of the reported score distribution and a corresponding increase in its quartiles.

Additionally, I find that nearly 25 percent of students in observed test-optional cohorts chose not to submit a college entrance exam score. There is significant variation across community; at treated New England Liberal Arts Colleges, nearly one in three enrolled students did not submit an SAT or ACT score. On the other hand, non-submitters make up just 7 and 11 percent, respectively, of test-optional cohorts at Elite National Universities and Liberal Arts Colleges in the Midwest, and treated members of Less Selective National Universities seem not to enroll any non-submitters. Estimates are precisely estimated, and I can reject at any conventional level that effects are equal across community.

Policy adoption also causes significant declines in the overall share of graduates majoring in a STEM field. There is substantial cross-community variation, with these declines concentrated in three communities; consequently, I can reject at the 5% level that effects on the Graduate STEM Share are equal across community. Declines are most precisely estimated for Colleges and Universities on the East Coast and Elite National Liberal Arts Colleges, with policy adoption causing a 16 and 8 percent decline, respectively, in the overall share of graduates majoring in a STEM field, with both point estimates significant at the 1% level. Those communities also see significant policy-induced declines of 26 and 23 percent, respectively, in the proportion of under-represented minority graduates majoring in a STEM field. These results are consistent with test-optional policies leading to a replacement of intermediate scorers with very low scorers and the literature on ability sorting across college majors.

There are still outstanding questions with respect to test-optional policies, especially regarding student outcomes and major choice. It is unclear what mechanism is driving the declines in the share of graduates with a STEM major. Non-submitters may prefer different majors; alternatively, it could reflect academic mismatch. Non-submitting students may *wish* to pursue a STEM major, but find themselves unprepared for the coursework; in the spirit of Arcidacono et al. (2011) and Arcidiacono et al. (2016), one could test for mismatch by comparing STEM major exit rates between submitting and non-submitting students.

Understanding the mechanism causing the decline has important implications for policymakers; if non-submitters would have completed a STEM major at a less selective institution, test-optional admission policies may have an adverse effect on the labor market outcomes of the students they aim to benefit, as many studies have documented wage premiums for STEM majors (see Grogger and Eide (1995), Loury (1997), Arcidiacono (2004), Kirkeboen et al. (2016)). Future research may also investigate the documented cross-community treatment effect heterogeneity; for instance, Elite National Liberal Arts Colleges may be less likely to consider applicant major when making admissions decisions, leading to a decline in STEM driven by compositional changes.

While the vast majority of selective institutions adopted a test-optional policy in response to the COVID-19 pandemic, several high profile institutions (such as MIT, Dartmouth, and Harvard) have reinstated their pre-pandemic testing policy. As more colleges and universities follow suit, investigating the impact of *rescinding* a test-optional policy would complement this analysis. The college admissions landscape may be vastly different in the post-pandemic world than in the pre-pandemic period I study. Of course, such an investigation would require many more colleges to return to a test-required policy; to the extent that colleges care about appearing both selective *and* diverse, my results suggest that there may be little incentive to abandon the new status-quo.

References

Arcidacono, P., Aucejo, E., Fang, H., and Spenner, K. (2011). Does affirmative action lead to mismatch? a new test and evidence. *Quantitative Economics*, 2(3):303–333.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1-2):343–375.

Arcidiacono, P., Aucejo, E. M., and Hotz, V. J. (2016). University differences in the graduation of minorities in stem fields: Evidence from california. *American Economic Review*, 106(3):525–562.

Belasco, A. S., Rosinger, K. O., and Hearn, J. C. (2015). The test-optional movement at america's selective liberal arts colleges: A boon for equity or something else? *Educational Evaluation and Policy Analysis*, 37(2):206–223.

Bennett, C. T. (2022). Untested admissions: Examining changes in application behaviors and student demographics under test-optional policies. *American Educational Research Journal*, 59(1):180–216.

Bevers, A. and Mulholland, S. E. (2022). Test-optional admissions and student debt. *Working Paper*.

Bollobas, B. (1980). A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, 1:311–316.

Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting event study designs: Robust and efficient estimation. *Review of Economic Studies*, 91:3253–3285.

Brandes, U., Delling, D., Gaertler, M., Gorke, R., Hoefer1, Martin Nikoloski, Z., and Wagner, D. (2008). On modularity clustering. *IEEE Transactions on Knowledge and Data Engineering 20*, 2:172–188.

Butts, K. and Gardner, J. (2021). *did2s: Two-Stage Difference-in-Differences Following Gardner (2021).*

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Conlin, M. and Dickert-Conlin, S. (2017). Inference by college admission departments. *Journal of Economic Behavior & Organization*, 141:14–28.

Conlin, M., Dickert-Conlin, S., and Chapman, G. (2013). Voluntary disclosure and the strategic behavior of colleges. *Journal of Economic Behavior & Organization*, 96:48–64.

Dessein, W., Frankel, A., and Kartik, N. (2025). Test-optional admissions. *American Economic Review*, forthcoming.

Fortunato, S. and Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports*, 659:1–44.

Furuta, J. (2017). Rationalization and student/school personhood in u.s. college admissions: The rise of test-optional policies, 1987 to 2015. *Sociology of Education*, 90:236–254.

Gardner, J., Thakral, N., Tô, L. T., and Yap, L. (2024). Two-stage differences in differences. *Working Paper*.

Goodman, J., Gurantz, O., and Smith, J. (2020). Take two! sat retaking and college enrollment gaps. *American Economic Journal: Economic Policy*, 12(2):115–158.

Goodman, S. (2016). Learning from the test: Raising selective college enrollment by providing information. *Review of Economics and Statistics*, 98(4):671–684.

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

Grogger, J. and Eide, E. (1995). Changes in college skills and the rise in the college wage premium. *Journal of Human Resources*, 30:280–310.

Hurwitz, M., Mbekeani, P. P., and Page, L. C. (2017). Surprising ripple effects: How changing the sat score-sending policy for low-income students impacts college access and success. *Educational Evaluation and Policy Analysis*, 39(1):77–103.

Hurwitz, M., Smith, J., Niu, S., and Howell, J. (2015). The maine question: How is 4-year college enrollment affected by mandatory college entrance exams? *Educational Evaluation and Policy Analysis*, 37:138–159.

Hyman, J. (2017). Act for all: The effect of mandatory college entrance exams on postsecondary attainment and choice. *Education Finance and Policy*, 12:281–311.

June, A. W. (2022). 'an art and a science': Colleges' tricky task of selecting peers. *The Chronicle of Higher Education*.

Kelly, J. (2022). Who benefits from multiple choice(s)?: The equilibrium impacts of test-optional college admissions. *Senior Essay. Yale University*.

Kirkeboen, L. J., Leuven, E., and Mogstad, M. (2016). Field of study, earnings, and self-selection. *Quarterly Journal of Economics*, 131(3):1057–1111.

Klasik, D. (2013). The act of enrollment: The college enrollment effects of state-required college entrance exam testing. *Education Researcher*, 42:151–160.

Knight, B. and Schiff, N. (2022). Reducing frictions in college admissions: Evidence from the common application. *American Economic Journal: Economic Policy*, 14(1):179–206.

Loury, L. (1997). The gender-earnings gap among college-educated worker. *Industrial and Labor Relations Review*, 50:580–593.

Luca, M. and Smith, J. (2013). Salience in quality disclosure: Evidence from the u.s. news college rankings. *Journal of Economics & Management Strategy*, 22(1):58–77.

McManus, B., Howell, J., and Hurwitz, M. (2023). Strategic disclosure of test scores: Evidence from us college admissions. *EdWorkingPaper*, 843.

National Center for Education Statistics (2022). Integrated postsecondary education data system. https://nces.ed.gov/ipeds.

Newman, M. E. J. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(026113):1–15.

Robinson, M. and Monks, J. (2005). Making sat scores optional in selective college admissions: a case study. *Economics of Education Review*, 24:393–405.

Rosenbaum, P. R. and Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1):33–38.

Rosinger, K. O. and Ford, K. S. (2019). Pell grant versus income data in postsecondary research. *Educational Researcher*, 48.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688.

Saboe, M. and Terrizzi, S. (2019). Sat optional policies: Do they influence graduate quality, selectivity or diversity? *Economics Letters*, 174:13–17.

Sirolly, A., Kanoria, Y., and Ma, H. (2024). The impact of race-blind and test-optional admissions on racial diversity and merit. *Working Paper*.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225:175–199.

Syverson, S. T., Franks, V. W., and Hicks, W. C. (2018). Defining access: How test-optional works. *National Association for College Admission Counseling*, pages 1–86.

The Chronicle of Higher Education (2020). Peer institutions. https://store.chronicle.com/products/peer-institutions.

Traag, V. A., Waltman, L., and van Eck, N. J. (2019). From louvain to leiden: guaranteeing well-connected communities. *Scientific Reports*, 9(5233):1–12.

Turner, S. and Bowen, W. (1999). Choice of major: the changing (unchanging) gender gap. *Industrial and Labor Relations Review*, 52:289–313.

Westrick, P. A., Marini, J. P., Young, L., Ng, H., and Shaw, E. J. (2023). Digital sat validity study - a comprehensive analysis of first-year college outcomes. *College Board Report*.

A Evaluating Community Strength

The Leiden algorithm will detect communities even when none are present (as in a randomly generated configuration network); therefore, it is natural to wonder if the detected communities are "real". Stating this more formally, do the detected communities lead to a higher value of modularity than would be expected from a realization of the configuration model? I follow the inference procedure described in Fortunato and Hric (2016) to determine if the observed modularity is significantly different from what would be expected if the peer institution network was randomly generated. Let $\mathcal{H}_O$ denote the observed modularity of the network given the communities identified by the Leiden algorithm. Let $G_{S,i}$ be a simulated network with the same nodes and node degree as the observed network, but with connections determined randomly by the configuration procedure. Let $\mathcal{H}_{S,i}$ denote the simulated modularity of $G_{S,i}$ given communities identified by applying the Leiden algorithm to the simulated network. After simulating 50 networks, I calculate

$$Z - Score = \frac{\mathcal{H}_O - \overline{\mathcal{H}_S}}{\sigma(\mathcal{H}_S)}, \tag{A.1}$$

where $\overline{\mathcal{H}_S}$ and $\sigma(\mathcal{H}_S)$ are the mean and standard deviation, respectively, of the simulated modularities. Table 1 displays the results of this analysis; the observed modularity is 18.34 standard deviations above the mean modularity in the simulated networks, which indicates that the peer institution network exhibits an incredibly strong community structure.

Table A1: Strength of Community Structure

$\mathcal{H}_O$	$\overline{\mathcal{H}}_S$	$\sigma(\mathcal{H}_S)$	$Z - Score$
0.819	0.499	0.0175	18.34

B Event-Study Tables

Table B1 presents the point estimates and standard errors used to construct Figures 6 through 10. The number in parentheses matches the number of the figure in the main text that the estimates correspond to. Coefficients represent average deviations from the estimated counterfactual potential outcome. Standard errors are from Gardner et al. (2024) and clustered at the institution level.

Table B2 presents treatment effect estimates and their standard errors when time fixed effects are estimated at the population-level rather than the community-level. There are severe pre-trends in many of the event-studies; for instance, pre-treatment effect sizes on the 25^{th} percentile of reported SAT scores (Table B2 Column 1) decrease monotonically from 6 to -14, and pre-treatment effect sizes on logged under-represented minority (URM) enrollment (Table B2 Column 3) rise almost monotonically from -0.03 to 0.07. These pre-trends severely limit ones ability to interpret post-treatment effect sizes and strongly support the use of community level time fixed effects when estimating treatment effects.

Table B1: Dynamic Effects of Policy Adoption on Outcome Variables

Outcome:	SAT 75^{th} %ile		SAT 25^{th} %ile	Test Submission Percent	Log Graduate STEM Share	
					Overall	URM-Only
Figure:	(6a)	(6b)	(7)	(8)	(9)	(10)
Variables						
T-$\geq$ 6	1.527	4.816***	2.401*	-0.7652*	0.0019	0.0471**
	(1.238)	(1.458)	(1.421)	(0.4002)	(0.0085)	(0.0203)
T-5	2.791	3.528*	3.248	-0.8452	0.0171	-0.0608
	(1.856)	(1.934)	(2.279)	(0.8846)	(0.0155)	(0.0511)
T-4	-1.113	-2.379	-0.9257	-0.1591	0.0023	0.0547
	(2.036)	(2.092)	(1.897)	(0.8750)	(0.0174)	(0.0561)
T-3	-2.743	-4.432**	-0.7151	0.9361	-0.0100	-0.1198**
	(1.956)	(2.172)	(2.230)	(0.8789)	(0.0184)	(0.0527)
T-2	-2.626	-7.680***	-3.502	1.226	0.0068	-0.0378
	(2.387)	(2.536)	(2.406)	(1.071)	(0.0163)	(0.0518)
T-1	-2.576	-9.329***	-8.408***	1.887	-0.0203	-0.0210
	(2.797)	(3.000)	(3.024)	(1.218)	(0.0201)	(0.0441)
T-0	4.162	-5.012	-1.196	-14.57***	-0.0490*	-0.0673
	(5.712)	(5.782)	(5.900)	(2.756)	(0.0275)	(0.0666)
T+1	13.23***	0.3769	19.24***	-17.96***	-0.0490*	-0.0760
	(4.557)	(4.076)	(6.377)	(2.878)	(0.0261)	(0.0757)
T+2	14.69***	1.768	22.10***	-19.94***	-0.0434*	0.0577
	(5.318)	(5.339)	(6.948)	(3.266)	(0.0259)	(0.0755)
T+3	12.42*	-0.9911	19.93***	-23.67***	-0.0622*	-0.0250
	(6.355)	(6.040)	(7.532)	(3.195)	(0.0345)	(0.0748)
Fit statistics						
Community Trends?	Yes	No	Yes	Yes	Yes	Yes
Observations	2,700	2,700	2,700	2,518	2,610	2,610
R^2	0.01656	0.00928	0.03128	0.22871	0.00929	0.00744
Adjusted R^2	0.01327	0.00597	0.02804	0.22594	0.00586	0.00400

Custom standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table B2: Dynamic Effects on All Variables - Population Trends

Outcome:	SAT 25th %ile	Test Submission Percent	Logged URM Enrlt.	Logged Enrlt. of Women	Logged Fed. Grant Recipient Enrlt.	Log Graduate Overall	STEM Share URM Only
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables							
T-$\geq$ 6	5.967***	-1.242**	-0.0264*	0.0168***	0.0065	0.0117	0.0851***
	(1.909)	(0.5403)	(0.0156)	(0.0053)	(0.0126)	(0.0103)	(0.0251)
T-5	3.490	-1.168	-0.0504*	0.0075	-0.0146	0.0006	-0.0870
	(2.646)	(0.8860)	(0.0304)	(0.0111)	(0.0190)	(0.0166)	(0.0549)
T-4	-2.086	-0.2493	0.0058	-0.0085	-0.0155	0.0095	0.0416
	(2.296)	(0.9006)	(0.0367)	(0.0101)	(0.0197)	(0.0181)	(0.0543)
T-3	-3.298	1.271	0.0310	-0.0217**	-0.0370	-0.0154	-0.1348**
	(2.323)	(0.9050)	(0.0291)	(0.0108)	(0.0262)	(0.0199)	(0.0562)
T-2	-9.458***	2.104	0.0481*	-0.0222*	-0.0064	-0.0067	-0.0887
	(2.732)	(1.309)	(0.0286)	(0.0118)	(0.0231)	(0.0180)	(0.0585)
T-1	-14.06***	2.983**	0.0700**	-0.0236*	0.0449*	-0.0313	-0.0678
	(3.301)	(1.317)	(0.0320)	(0.0140)	(0.0231)	(0.0234)	(0.0493)
T-0	-10.96**	-13.15***	0.1465***	0.0037	0.0949***	-0.0571**	-0.0929
	(5.549)	(2.478)	(0.0443)	(0.0173)	(0.0320)	(0.0277)	(0.0581)
T+1	6.481	-16.04***	0.1548***	-0.0031	0.0719**	-0.0666***	-0.1235*
	(5.900)	(2.635)	(0.0491)	(0.0197)	(0.0319)	(0.0243)	(0.0649)
T+2	6.849	-17.46***	0.2034***	-0.0157	0.0802**	-0.0618**	-0.0135
	(6.839)	(2.954)	(0.0446)	(0.0223)	(0.0342)	(0.0260)	(0.0587)
T+3	5.118	-21.21***	0.1684***	-0.0252	0.0965**	-0.0971***	-0.1044*
	(7.078)	(2.929)	(0.0504)	(0.0231)	(0.0471)	(0.0325)	(0.0606)
Fit statistics							
Community Trends?	No	No	No	No	No	No	No
Observations	2,700	2,518	2,872	2,874	2,873	2,610	2,610
R^2	0.01619	0.18918	0.03628	0.00568	0.01298	0.01684	0.01575
Adjusted R^2	0.01290	0.18627	0.03325	0.00256	0.00988	0.01344	0.01235

Custom standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

C Alternative Classification

In this appendix, I consider an alternative (simpler) classification of institutions that relies on their Carnegie Classification rather than the peer network. First, I classify institutions along two dimensions: degree-granting level and selectivity. These are the two dimensions considered (separately) by Bennett (2022) in his heterogeneity analysis. First, I distinguish between institutions whose highest degree offered is a bachelor's degree and those that also offer graduate degrees (master's or doctorates). Second, I divide institutions into more selective and less selective groups based on their undergraduate profile; institutions rated "More Selective" by the Carnegie Classification are in the more selective category, and all other institutions are considered less selective. This results in four institutional categories: (1) less selective, bachelor's-only institutions; (2) more selective, bachelor's-only institutions; (3) less selective, graduate-granting institutions; and (4) more selective, graduate-granting institutions. Table B1 describes these four groups and the number of treated and never-treated institutions in each.

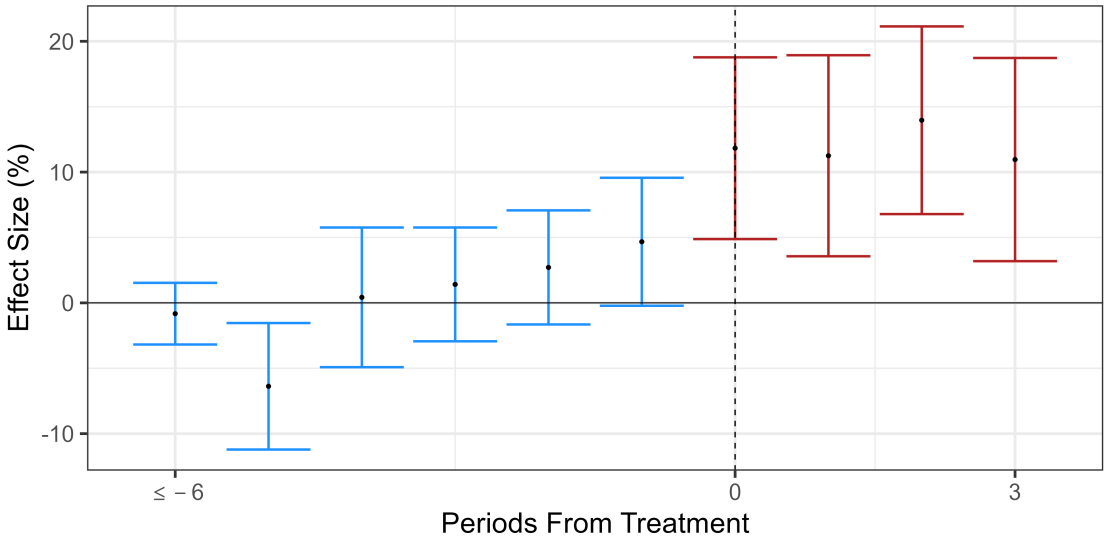
Table C1: Groups based on Carnegie Classification

Group	Treated	Never-Treated	Median Treatment Year
Less Selective, Bachelor's-Only	22	30	2010.50
More Selective, Bachelor's-Only	13	28	2010
Less Selective, Graduate-Granting	29	71	2012
More Selective, Graduate-Granting	8	49	2012

Note: Only communities with 2 or more treated institutions are included. An institution is treated if it adopted a test-optional policy between 2006 and 2016, and Never-Treated if it did not adopt a test-optional policy by 2020. Institutions that adopted a policy between 2017 and 2019 are excluded.

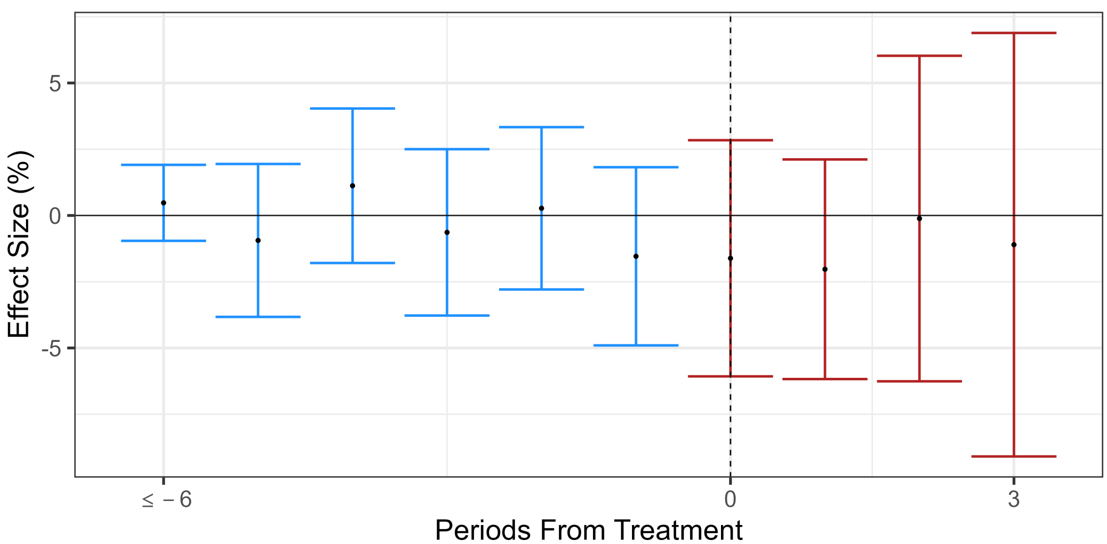
I repeat the empirical analysis for dynamic effects with group-specific trends corresponding to the aforementioned groups. Table C2 presents the event study estimates for the eight outcomes of interest. Figures C1 and C2 show plots of the event study estimates for logged FTFT URM enrollment and the logged Graduate STEM Share. Figure C1 has a clear positive pre-trend that is not present when community trends are based on network detected communities (Appendix Figure D1). Post-treatment estimates are larger, likely biased upwards by the positive pre-trend and possibly overstating the effect of policy adoption on improving racial diversity. In contrast to Figure 12 in the main text, Figure C2 shows no effect of policy adoption on the logged Graduate STEM Share.

Figure C1: Effect of Policy Adoption on Logged FTFT URM Enrollment - Carnegie Classification Trends



Note: Estimates from model with carnegie classification group-specific trends. FTFT = First-Time Full-Time. URM = under-represented minority. Coefficients represent deviations from the counterfactual potential outcome. Standard errors are from Gardner et al. (2024) and clustered at the institution level. Error bars represent 90 percent confidence intervals. Point estimates found in Table C2.

Figure C2: Effect of Policy Adoption on Log Graduate STEM Share - Carnegie Classification Trends



Note: Estimates from model with carnegie classification group-specific trends. Coefficients represent deviations from the counterfactual potential outcome. Standard errors are from Gardner et al. (2024) and clustered at the institution level. Error bars represent 90 percent confidence intervals. Point estimates found in Table C2.

Table C2: Dynamic Effects on All Variables - Carnegie Classification Trends

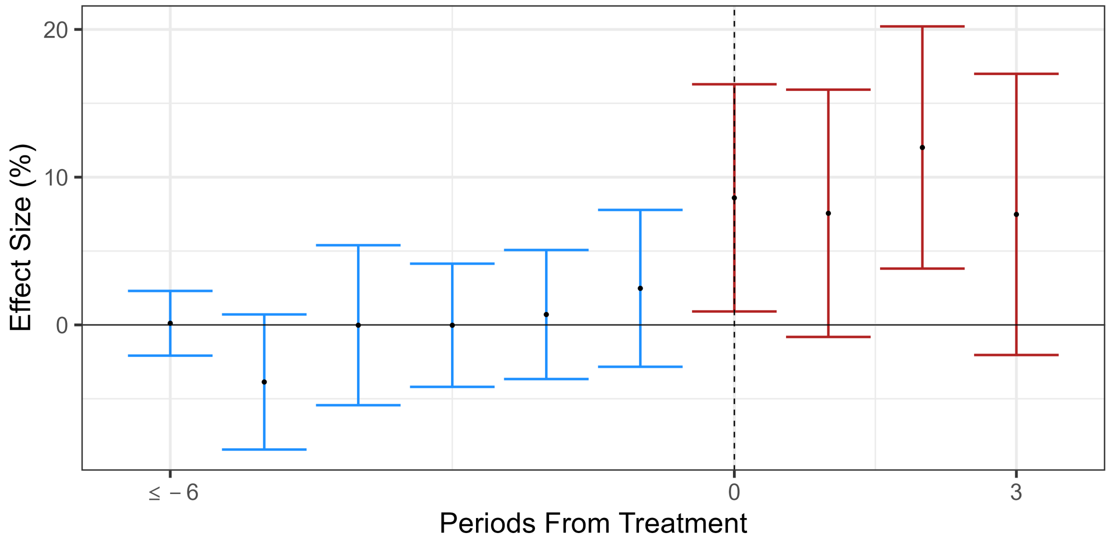
Outcome:	SAT 25th %ile	SAT 25th %ile	Test Submission Percent	Logged URM Enrlt.	Logged Enrlt. of Women	Logged Fed. Grant Recipient Enrlt.	Log Graduate STEM Share Overall	Log Graduate STEM Share URM Only
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables								
T-$\geq$ 6	1.812	0.8288	-1.029**	-0.0083	0.0087	0.0101	0.0048	0.0644***
	(1.448)	(1.243)	(0.4363)	(0.0143)	(0.0066)	(0.0108)	(0.0087)	(0.0222)
T-5	3.182	2.903	-0.6955	-0.0638**	0.0056	-0.0287	-0.0094	-0.0878*
	(2.488)	(1.804)	(0.9259)	(0.0293)	(0.0102)	(0.0191)	(0.0175)	(0.0500)
T-4	-0.5067	-0.8796	-0.5909	0.0042	-0.0021	-0.0225	0.0112	0.0385
	(1.974)	(2.074)	(0.8610)	(0.0324)	(0.0103)	(0.0180)	(0.0177)	(0.0491)
T-3	0.4936	-1.045	1.108	0.0141	-0.0153	-0.0295	-0.0064	-0.0968*
	(2.080)	(1.933)	(0.8775)	(0.0264)	(0.0109)	(0.0242)	(0.0190)	(0.0530)
T-2	-4.193*	-2.470	1.757	0.0271	-0.0140	-0.0061	0.0027	-0.0737
	(2.296)	(2.320)	(1.117)	(0.0264)	(0.0119)	(0.0207)	(0.0186)	(0.0546)
T-1	-6.489**	-1.816	2.354**	0.0467	-0.0088	0.0427**	-0.0154	-0.0291
	(2.759)	(2.525)	(1.136)	(0.0297)	(0.0134)	(0.0212)	(0.0204)	(0.0435)
T-0	-1.119	5.241	-13.23***	0.1183***	0.0170	0.0765**	-0.0162	-0.0046
	(5.332)	(5.472)	(2.392)	(0.0421)	(0.0175)	(0.0310)	(0.0270)	(0.0603)
T+1	15.50***	10.38**	-16.54***	0.1125**	0.0191	0.0595**	-0.0203	-0.0642
	(6.001)	(4.394)	(2.528)	(0.0466)	(0.0196)	(0.0299)	(0.0251)	(0.0685)
T+2	18.18***	13.08***	-17.89***	0.1396***	0.0059	0.0506	-0.0012	0.0837
	(6.675)	(4.926)	(2.874)	(0.0435)	(0.0232)	(0.0335)	(0.0372)	(0.0708)
T+3	17.06**	12.55**	-21.64***	0.1096**	-0.0024	0.0722*	-0.0110	0.0064
	(6.969)	(5.794)	(2.705)	(0.0471)	(0.0241)	(0.0425)	(0.0484)	(0.0767)
Fit statistics								
Community Trends?	Based on Carnegie Classification Groups							
Observations	3,028	3,028	2,837	3,237	3,237	3,236	2,947	2,947
R^2	0.02044	0.01184	0.19885	0.01991	0.00194	0.00818	0.00056	0.00723
Adjusted R^2	0.01751	0.00889	0.19630	0.01718	-0.00084	0.00541	-0.00251	0.00419

Custom standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

D Effects on Student-Body Diversity

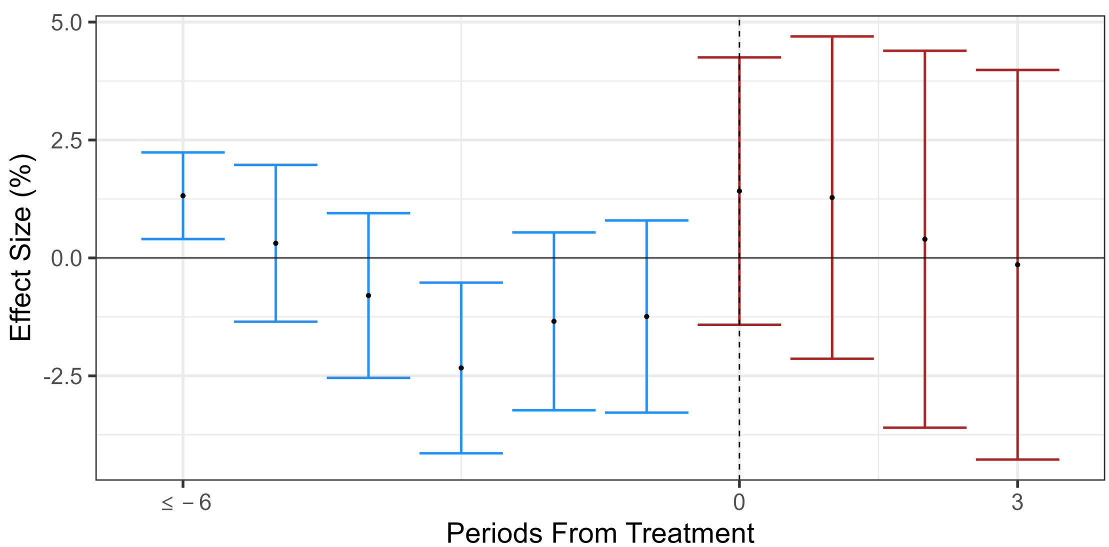
D.1 Event-Studies

Figure D1: Effect of Policy Adoption on Logged FTFT URM Enrollment



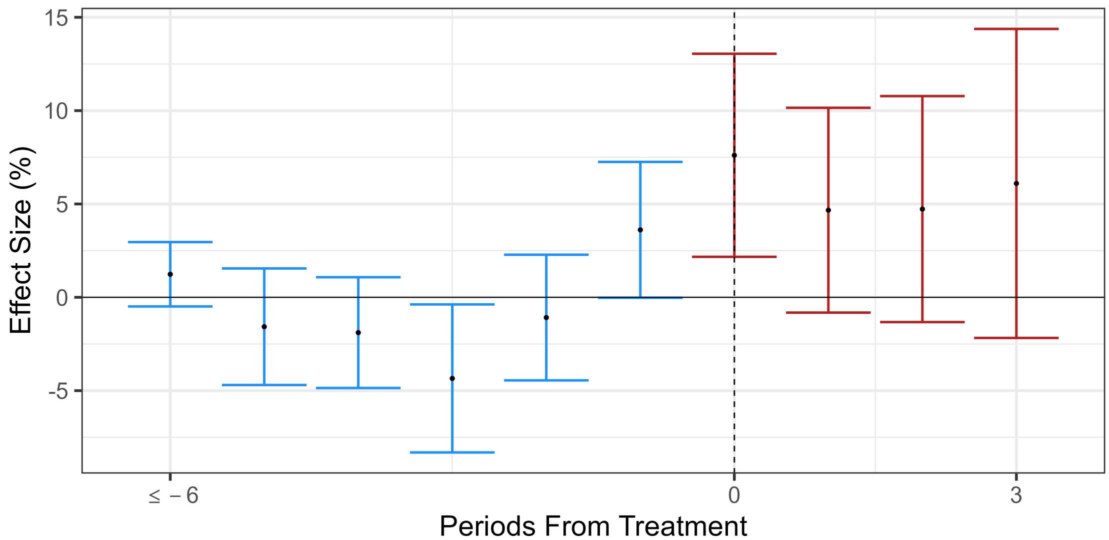
Note: FTFT = First-time full-time. URM = Under-Represented Minority. Estimates from model with community-specific trends. Coefficients represent deviations from the counterfactual potential outcome. Standard errors are from Gardner et al. (2024) and clustered at the institution level. Error bars represent 90 percent confidence intervals. Point estimates found in Appendix Table D2.

Figure D2: Effect of Policy Adoption on Logged FTFT Enrollment of Women



Note: FTFT = First-time full-time. Estimates from model with community-specific trends. Coefficients represent deviations from the counterfactual potential outcome. Standard errors are from Gardner et al. (2024) and clustered at the institution level. Error bars represent 90 percent confidence intervals. Point estimates found in Appendix Table D2.

Figure D3: Effect of Policy Adoption on Logged FTFT Enrollment of Fed. Grant Recipients



Note: FTFT = First-time full-time. Estimates from model with community-specific trends. Coefficients represent deviations from the counterfactual potential outcome. Standard errors are from Gardner et al. (2024) and clustered at the institution level. Error bars represent 90 percent confidence intervals. Point estimates found in Appendix Table D2.

Figures D1, D2, and D3 present dynamic effects of policy adoption on logged first-time full-time (FTFT) enrollment of under-represented minorities (URMs), women, and federal grant recipients, respectively. My estimates largely agree with the findings of Bennett (2022), though my point estimates are, in general, slightly smaller (not statistically significantly so). At treatment onset, effects on FTFT URM enrollment increase sharply and are almost 9 percent higher than in the never-treated counterfactual. Effects stay positive thereafter, and are statistically significant in two of the four post-treatment periods. Effects on the enrollment of women are smaller. At treatment onset, FTFT enrollment of women is around 1.3 percent higher than would be expected in the absence of treatment, and estimates decline thereafter; however, estimates would be about twice as large if compared to the last pre-treatment period, though they would still not be statistically significant. Effects on the enrollment of federal grant recipients fall somewhere between the effects for under-represented minorities and women. At treatment onset, FTFT federal grant recipient enrollment is about 7.6 percent higher than in the never-treated counterfactual, and effect sizes are positive in all post-treatment periods; however, the last pre-treatment estimate is around 0.036. Thus, the estimated effect would be smaller if the post-treatment comparison was the T-1 estimate. Point estimates can be found in Table D2.

D.2 Community-Level Treatment Effects

Table D1: Effects on Student Demographics by Community

Outcome:	Logged URM Enrlt. (1)	Logged Enrlt. of Women (2)	Logged Fed. Grant Recipient Enrlt. (3)
Community			
Less Selective Nat'l Unis	-0.0951	-0.1236***	-0.0795
	(0.0871)	(0.0425)	(0.0600)
LACs - New England	0.2122	-0.0295	0.0454
	(0.2040)	(0.0717)	(0.1032)
Colleges and Unis - East	0.0019	0.0456	0.0790
	(0.0797)	(0.0350)	(0.0659)
LACs - Midwest	0.0433	0.0029	0.1011*
	(0.0738)	(0.0337)	(0.0541)
Colleges and Unis - Midwest	-0.0394	-0.0631	-0.0308
	(0.1262)	(0.0722)	(0.1098)
Colleges and Unis - Southeast	0.1424	0.0137	-0.0980
	(0.1150)	(0.0816)	(0.1218)
Elite Nat'l LACs	0.0853	0.0126	-0.0213
	(0.0603)	(0.0157)	(0.0629)
Elite Nat'l Unis	0.2190***	0.0954***	0.2634**
	(0.0492)	(0.0293)	(0.1142)
Prob. All Effects Equal	0.017	<0.01	<0.01
Fit statistics			
Observations	2,988	2,990	3,099
R^2	0.01886	0.01406	0.01447
Adjusted R^2	0.01656	0.01175	0.01224

Custom standard-errors in parentheses, clustered at institution-level.
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table D1 presents community-level static treatment effect estimates for the three diversity outcomes. Elite National Universities see statistically significant increases in all three outcomes; however, a pre-trend investigation reveals that – despite mostly stable pre-treatment estimates – the final pre-treatment estimate for each outcome is large and positive. This suggests that other policies may have been enacted prior to test-optional admissions with the goal of increasing student-body diversity.

Enrollment of federal grant recipients increases by 10, 8, and 5 percent at Liberal Arts Colleges - Midwest, Colleges and Universities - East, and liberal Arts Colleges - New England, but only the first of these estimates is statistically significant, and none are statistically

Table D2: Dynamic Effects of Policy Adoption on Student-Body Diversity

Outcome:	Logged URM Enrlt.	Logged Enrlt. of Women	Logged Fed. Grant Recipient Enrlt.
Model:	(9)	(10)	(11)
Variables			
T-$\geq$ 6	0.0011	0.0132**	0.0123
	(0.0133)	(0.0055)	(0.0105)
T-5	-0.0386	0.0031	-0.0158
	(0.0278)	(0.0100)	(0.0187)
T-4	-0.0002	-0.0080	-0.0189
	(0.0326)	(0.0106)	(0.0181)
T-3	-0.0002	-0.0233**	-0.0435*
	(0.0255)	(0.0110)	(0.0239)
T-2	0.0070	-0.0134	-0.0108
	(0.0266)	(0.0111)	(0.0205)
T-1	0.0248	-0.0124	0.0361
	(0.0321)	(0.0123)	(0.0221)
T-0	0.0860*	0.0142	0.0761**
	(0.0468)	(0.0172)	(0.0331)
T+1	0.0756	0.0128	0.0467
	(0.0509)	(0.0207)	(0.0332)
T+2	0.1201**	0.0040	0.0473
	(0.0496)	(0.0239)	(0.0366)
T+3	0.0748	-0.0014	0.0610
	(0.0582)	(0.0248)	(0.0501)
Fit statistics			
Community Trends?	Yes	Yes	Yes
Observations	2,872	2,874	2,873
R^2	0.01124	0.00335	0.00739
Adjusted R^2	0.00813	0.00021	0.00427

Custom standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

significant from the others. The other communities see imprecisely estimated declines in federal grant recipient enrollment.

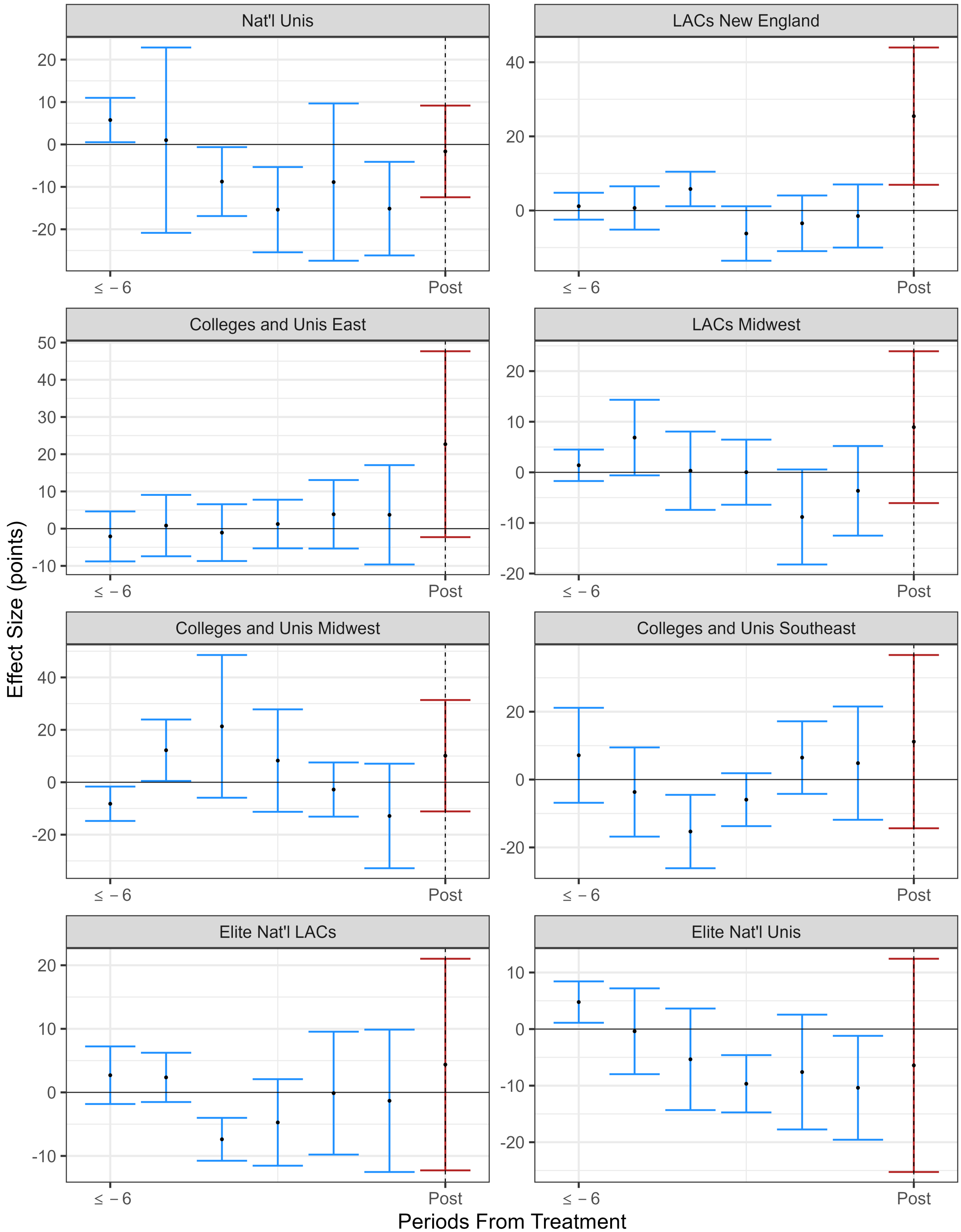
E Community-Level Event Studies

In this appendix, I report estimates from community-by-community regressions of deviations from counterfactual potential outcomes (d_{ct}) on treatment leads plus a post-treatment indicator. Treatment effects estimated in the pre-period will allow me to determine if treated and control units in the same community evolved similarly prior to treatment.

Figure E1 displays these community-level event studies for the 75$^{\text{th}}$ percentile of reported SAT scores. Pre-treatment estimates are around zero for most communities except for Less Selective National Universities (top left panel) and Elite National Universities (bottom right panel). These communities exhibit strong pre-trends, suggesting that treated members of these communities were becoming less selective (relative to untreated members) prior to treatment. Figure E2 displays community-level event studies for the 25$^{\text{th}}$ percentile, where a qualitatively similar but quantitatively more extreme pattern emerges. These negative trends are partly responsible for the negative pre-trend in Figure 7. These pre-trends suggest that reported SAT scores were increasing faster at untreated members of these communities than treated members. Thus, for this outcome in particular, the assumption of parallel counterfactual trends may not hold for those community/outcome combinations specifically.

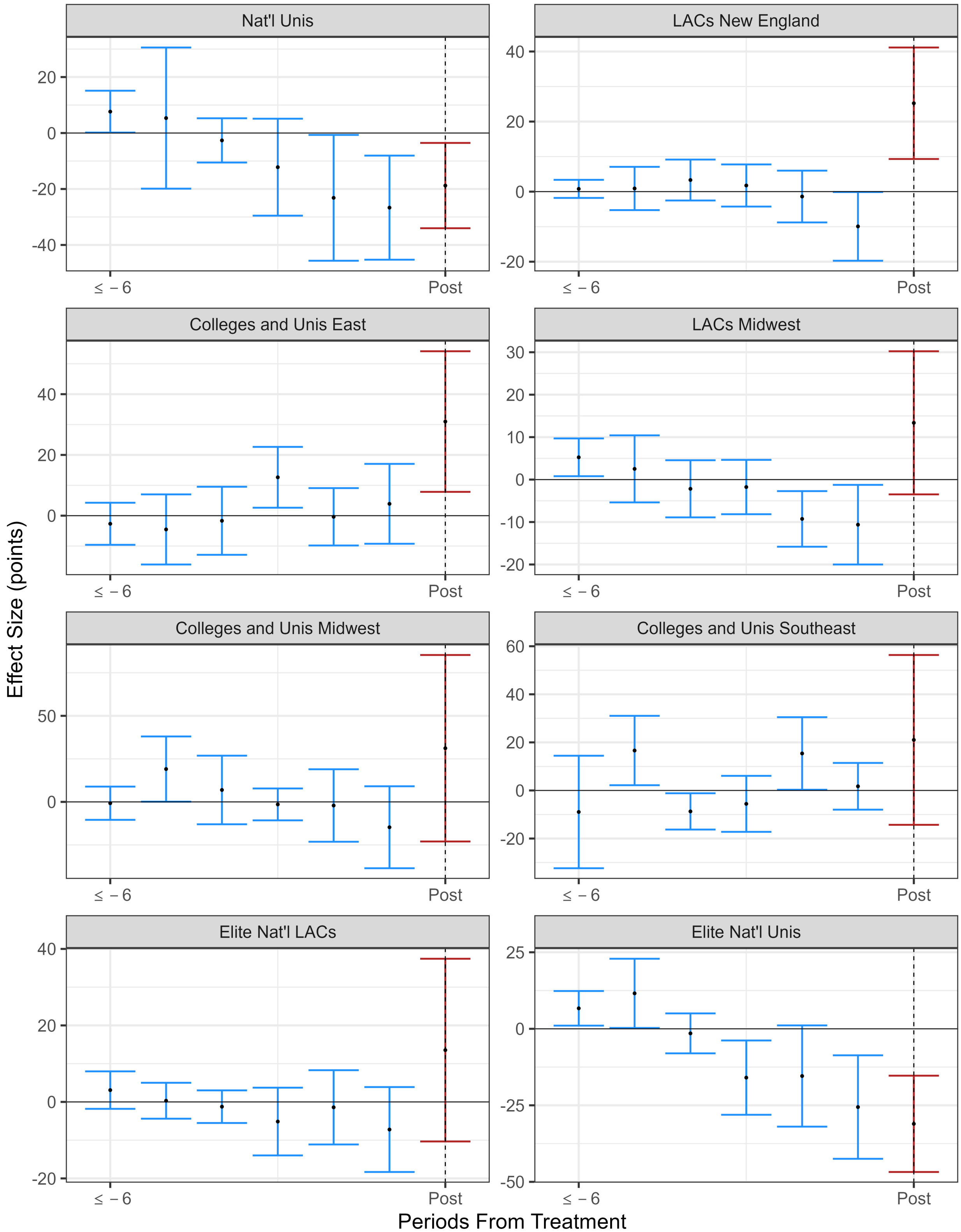
Figure E3 displays community-level event studies for the overall Graduate STEM Share. It is clear that the large, statistically significant declines at Colleges and Universities on the East Coast and Elite National Liberal Arts Colleges are not driven by different underlying trends between treated and control units in those communities. Pre-treatment estimates for both are stable and drop sharply post-treatment. Figure E4 tells a similar story for the URM-only Graduate STEM Share in these communities, with stable pre-trends followed by a drop at treatment onset.

Figure E1: Community-Level Effects on 75th %ile of Reported SAT Scores



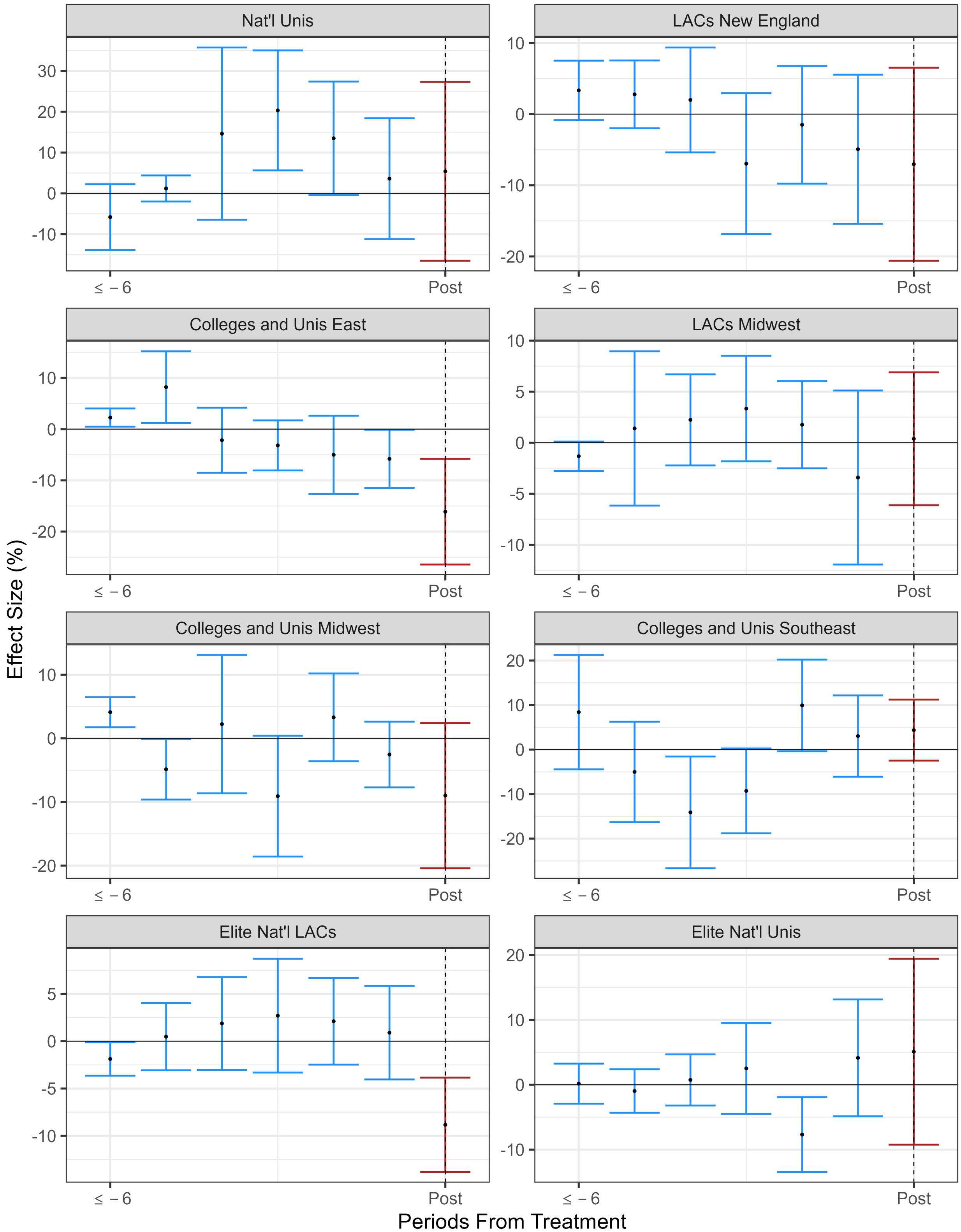
Note: Coefficients represent deviations from the counterfactual potential outcome. Standard errors are from Gardner et al. (2024) and clustered at the institution level. Error bars represent 90 percent confidence intervals.

Figure E2: Community-Level Effects on 25th %ile of Reported SAT Scores



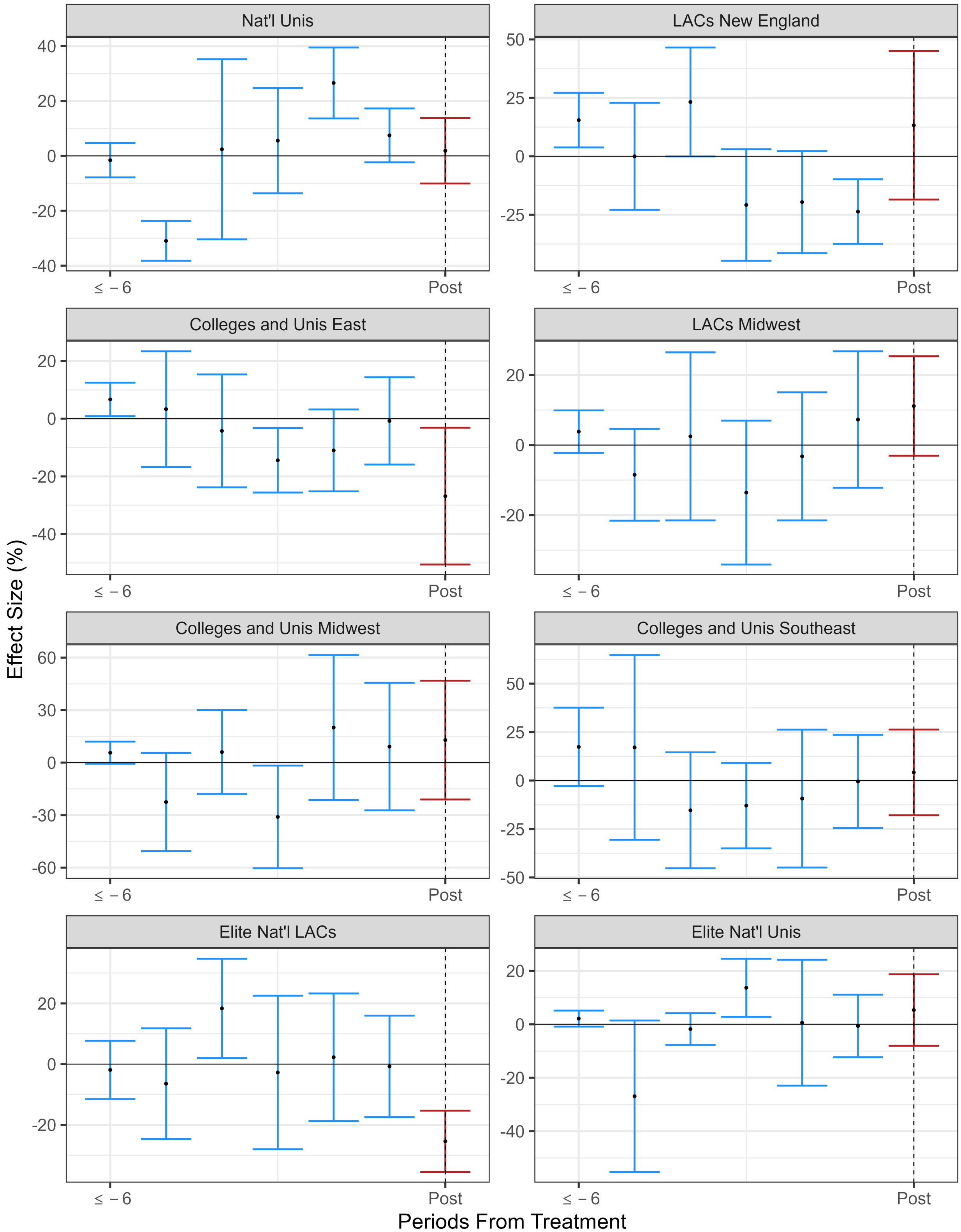
Note: Coefficients represent deviations from the counterfactual potential outcome. Standard errors are from Gardner et al. (2024) and clustered at the institution level. Error bars represent 90 percent confidence intervals.

Figure E3: Community-Level Effects on Overall Graduate STEM Share



Note: Coefficients represent deviations from the counterfactual potential outcome. Standard errors are from Gardner et al. (2024) and clustered at the institution level. Error bars represent 90 percent confidence intervals.

Figure E4: Community-Level Effects on URM-Only Graduate STEM Share



Note: Coefficients represent deviations from the counterfactual potential outcome. Standard errors are from Gardner et al. (2024) and clustered at the institution level. Error bars represent 90 percent confidence intervals.