Disrupting Education's Old Guard: The Labor Market Effects of Replacing Traditional Class Exams with Group Projects*

Alonso Alfaro-Ureña†

Marco A. Badilla Maroto‡

September 22, 2025

Abstract

To promote deeper learning and fairer outcomes, many education systems have moved from traditional exams to lower-stakes assessments, yet evidence remains limited. This study evaluates a 2008 reform at ULACIT University in Costa Rica that replaced all class exams with group projects. Using a differences-in-differences design, we find sizable labor market gains, including higher wages and formal employment. These effects partly reflect greater work experience during college and enhanced soft skills, especially networking. Gains are strongest for men, public high school graduates, and lower performers. The reform also spurred grade inflation, faculty turnover, and slower enrollment growth.

*We thank Sydnee Caldwell, Benjamin Faber, Jesse Rothstein, and Christopher Walters for incredible guidance and support. We also thank Hilary Hoynes, Patrick Kline and Valeria Lentini, and participants in the UC Berkeley, Uppsala University and University of Costa Rica's Economics Seminars for helpful comments and feedback. We acknowledge funding provided by The Center for Latin American and Caribbean Studies (CLACS) at UC Berkeley. The views expressed herein are solely of the authors and are not necessarily those of the Central Bank of Costa Rica (BCCR). All results have been reviewed to ensure that no confidential information is disclosed. All remaining errors are our own.

†Banco Central de Costa Rica (BCCR) and Universidad de Costa Rica (UCR). Email: alfaroua@bccr.fi.cr

‡UC Berkeley. Email: mbadilla@berkeley.edu

1. Introduction

Exams remain the dominant mode of classroom evaluation in higher education, typically accounting for over half of students' course grades.[1] Yet an increasing number of institutions are shifting away from high-stakes exams, replacing them with alternatives such as open-book tests, take-home assignments, homework graded for completion, and opportunities for retakes—often with the goal of improving outcomes for underrepresented groups.[2] Despite this growing shift, little is known about the broader consequences of abandoning exam-centered evaluation. This gap is critical because long-term returns to education appear to operate more through the development of non-cognitive skills than through test-based knowledge (Jackson 2018).

This paper provides new evidence on the consequences of eliminating exams as a method of classroom evaluation. We study the 2008 reform at ULACIT, a highly ranked university in Costa Rica, which abolished exams and replaced them with group projects across all classes and programs. This reform offers a rare opportunity to assess how such a sweeping change in evaluation methods affects student composition, faculty behavior, and graduates' labor market outcomes. We also examine the mechanisms behind these effects and their heterogeneity by gender, family income, and student ability. To our knowledge, this is the first study to analyze the complete removal of exams in an education setting.

To answer this question, we combine administrative labor microdata from Social Security records, college graduation data from the country's major universities, individual high school academic and demographic information from the Ministry of Education, and family networks from the Civil Registry. We begin by using university-level statistics to document that the policy led to shorter program durations—due to the more flexible nature of project-based assessments—as well as grade inflation, driven by their lower-stakes nature.

To causally estimate the effect of this policy, we implement a difference-in-differences (DiD) design that leverages the timing of the reform and compares treated admission cohorts at

1. Based on a sample of 500 syllabi from multiple U.S. universities, 50–60% of course grades are determined by exams (quizzes, midterms, finals). This estimation is computed using 500 syllabi, from all academic areas, from the online open repositories of the following U.S. universities: MIT, San Diego State University, University of Michigan, University of Connecticut, University of Florida, UC Berkeley, and Yale.

2. For recent global evidence of such reforms, see Johnston (2025) on Cambridge and Oxford, and Tucker (2025) on the San Francisco Unified School District.

ULACIT to cohorts at other universities of comparable academic ranking. Given the idiosyncratic nature of the reform—being specific to a single institution—we investigate the possibility that the policy was bundled with or correlated to other simultaneous reforms at ULACIT. To address this concern, we examine archival records from the Ministry of Education, which documents all policy and curriculum changes across private universities in the country. We construct a detailed timeline of major reforms at ULACIT and find no confounding changes around the time of interest, aside from an infrastructure project that took place several years later. All pre-2008 changes occurred far enough in the past to assume that any potential effects were stable by 2008 and can be captured through the DiD method as part of the university selection bias. We further show that our results are not driven by the specific majors that benefited from the mentioned infrastructure project.

To structure our empirical investigation, we develop a conceptual framework that groups the effects of the reform into three primary domains: changes in student composition, faculty responses, and graduates' labor market outcomes. First, the policy may have altered the type of students enrolling at ULACIT—whether by attracting students who prefer collaborative learning or discouraging those who value traditional exams. These compositional shifts are observable in characteristics such as academic ability and socioeconomic background. Because the policy was enacted after several cohorts had already enrolled, we can cleanly separate its causal effects from post-reform changes in student composition. This distinction is central to our identification strategy and is made explicit throughout the analysis. While dropout data are not available, we later provide indirect evidence that attrition did not increase for the partially treated cohorts, minimizing concerns about survivor bias in this comparison.

Second, the reform may have influenced faculty behavior, particularly through changes in workload, grading practices, and pedagogical alignment. Instructors opposed to the new system or overwhelmed by the increased demands may have exited the university, leading to shifts in teacher retention and compensation. Third, we analyze a range of labor market outcomes—including formal employment and wages—and interpret them through several suggestive reinforcing mechanisms. These include work experience during college, the development of soft skills such as communication, leadership, and networking; signaling effects, whereby grade compression or changes in transcript credibility may grant graduates access to better-paying firms through camouflage; occupational re-sorting into jobs more aligned with

collaborative training; and potential learning gains. Together, these mechanisms offer a suggestive lens to interpret how a shift in evaluation methods can shape long-run academic and labor market trajectories.

It is important to emphasize that we interpret all of these domains not as independent shocks coinciding with the reform, but as endogenous responses to it. In other words, the "effect of the reform" should be understood as the total bundle of consequences triggered by the elimination of exams—from shifts in student composition and faculty turnover to changes in the evaluation practices. Our main labor outcome estimates therefore capture the reform's overall impact on graduates, while complementary analyses help to disentangle the specific channels through which this impact operated.

We begin our analysis by examining changes in student composition following the reform. Post-2008 cohorts, exhibit a 4.2% decline in total enrollment at ULACIT relative to the control group. This decline appears to reflect hesitation among prospective students facing an unfamiliar and unconventional evaluation system. However, we also observe a 9.3 percentage point increase in graduate enrollment as a share of total enrollment, suggesting that graduate programs—already less reliant on traditional exams—were more compatible with the new assessment model. Contrary to institutional concerns that the policy would attract lower-performing students, we show that math proficiency among incoming students actually increased by one point (out of a 100 scale) after the policy change. We detect no significant change in gender composition, but do find a 10 percentage point increase in the share of students from private high schools, as well as a temporary decline in the enrollment of students whose parents held a college degree.

The reform had important consequences for faculty. Turnover increased by 7.3 percentage points, as some instructors opposed the new pedagogical model or found the added workload unsustainable. Leavers were disproportionately part-time, less likely to be ULACIT alumni, and less likely to hold graduate degrees—suggesting the policy did not drive out the most experienced faculty, but rather those with weaker institutional ties and less alignment with project-based instruction. While this introduced short-term instability, it also enabled a transition toward faculty better suited to the new system. At the same time, stagnant wages—driven in part by slower enrollment—widened salary gaps relative to peer institutions, compounding long-term recruitment and retention challenges.

We also find suggestive evidence of improved soft skills, particularly in social connectedness: ULACIT graduates experienced a 6.5 percentage point increase in the number of college peers among their co-workers, measured relative to the size of their admission cohort within the same major, which may reflect the collaborative nature of project-based evaluations. However, this growth in alumni networks did not correspond to a rise in managerial positions, suggesting that while the reform strengthened peer relationships, it did not directly foster leadership development.

The reform led to substantial labor market improvements. ULACIT students became 13.8 percentage points more likely to work during college and 8.3 percentage points more likely to be formally employed after graduation—consistent with better alignment between coursework and real-world job demands. Graduates also experienced a 18% relative wage gain, with effects robust to alternative specifications, including analyses restricted to the private university sample, placebo tests reassigning pseudo-treatment to control universities, and estimations excluding control graduates potentially exposed to spillovers. A decomposition exercise suggests that approximately 29% of the wage premium may be accounted by greater work experience during college, 20% by stronger alumni networks, 12% by access to higher-paying firms, and 6% by occupational re-sorting. The remaining 33% could be consistent with unmeasured gains in learning or other soft skills, and should be viewed as capturing other factors we cannot directly observe. While our results are not a call to eliminate exams, they underscore that competencies under-incentivized in exam-centered systems are materially rewarded in the labor market. An additional 20 percentage point wage gain emerges among post-2008 cohorts, driven by compositional shifts toward higher-ability students, greater private school representation, and increased graduate program enrollment.

The wage gains from the reform are concentrated among undergraduates, with no effect among graduate students—who already operated under similar evaluation systems—serving as a robust placebo test. The gains are significantly larger for male students, despite female students benefiting more from work experience during college and alumni networks, indicating that unexplained factors play a role. Effects are strongest among students from public high schools, who appear to have gained greater work experience during college and access to better-paying firms—perhaps because they had more to gain from improved career alignment than their private counterparts. Similarly, students with below-median high school GPA ben-

efited disproportionately, which may partly reflect grade compression and the signaling value of academic camouflage helping them overcome weaker academic records.

As for external validity, several of the mechanisms we document—such as soft skill development and occupational re-sorting—could plausibly extend to higher education systems in countries like the U.S., though wage gains may be smaller given the already existing work exposure at the college level. Importantly, we speculate that many of the reform's strongest effects hinge on ULACIT's early adopter status. A nationwide rollout—where no university stands out for implementing such a policy—could plausibly weaken or eliminate several of the advantages we observe. Finally, the policy's lack of adoption at other universities reflects explicit institutional confidence in exam-based systems and their perceived value in measuring individual ability—a view confirmed through discussions with university administrators and policymakers—despite recognition of the results observed at ULACIT.

This paper contributes to three strands of literature. First, it adds to the literature on formative assessment, which emphasizes the benefits of frequent, lower-stakes evaluations and increased feedback. Existing studies demonstrate the positive effects of homework on test scores and college attendance (Eren and Henderson (2008), Eren and Henderson (2011), Grodner and Rupp (2013) and Kalenkoskia and Pabilonia (2016)) and highlight the benefits of project-based learning (Kokotsaki et al. (2016), Kingston (2018), Saavedra et al. (2022) and Krajcik et al. (2023)). Additionally, Berry et al. (2020) explores how increasing evaluation frequency affects student performance. Our main contribution to this literature lies in examining labor market outcomes, an area that remains underexplored in prior studies. Unlike previous work, which focuses primarily on K–12 education, we analyze formative assessment policies in higher education, where results may differ due to variations in student characteristics and connections to the labor market. Furthermore, this paper examines a unique policy that fully replaces exams with group projects; in traditional systems, projects are typically used as a supplement to exams, often leading to challenges like low effort and free-riding. By substituting exams entirely with projects, this policy introduces new dynamics, including shifts in effort allocation and the potential for reputation-based incentives to mitigate free-riding behavior.[3]

Second, this paper contributes to the broader literature on the role of exams in education.

3. Importantly, students engage in group projects within each class, across classes, and throughout the program, often working with peers from their admission cohort. Groups are self-selected by students, influencing teamwork and peer interaction.

While much attention has focused on standardized admission exams to college,[4] this literature also examines the signaling value of other exams (Clark and Martorell (2014), Schwerdt and Woessmann (2017), Diamond and Persson (2017)) and their drawbacks, such as encouraging memorization and increasing anxiety (Ebenstein et al. (2016), Heissel et al. (2021) and Lincove et al. (2022)). By investigating the complete removal of exams as class evaluative tools within a university setting, this paper challenges the traditional role of exams and explores their implications for both academic and labor market outcomes.

Third, we contribute to the growing literature on the importance of soft skills in labor market outcomes. Studies such as Deming (2017) highlight how different skills, like communication (Deming and Kahn 2018), teamwork (Deming and Weidmann 2021), and social or alumni networks (Fischer et al. 2023), are increasingly valued by employers and contribute to career success. Jackson (2018) provides further evidence that teacher effects on long-term student success operate largely through non-test-score channels, reinforcing the idea that alternative evaluation systems—such as project-based learning—may cultivate the kinds of skills that drive labor market success. ULACIT's policy of replacing exams with group projects provides a natural experiment to assess how fostering collaborative learning can enhance networking, leadership, teamwork, and communication skills, and subsequently affect students' labor market success.

The rest of the paper is organized as follows. Section 2 provides the setting, section 3 describes the data, section 4 details the exam-to-projects policy, section 5 lays out the framework used to interpret the policy's effects, section 6 explains the empirical strategy, section 7 presents and discusses the results and finally section 8 concludes.

2. Setting

To address our research question, we examine a natural experiment at the university level in Costa Rica, a high-income developing country in Central America with a population of approximately 5 million people. Costa Rica's higher education system consists of public and private

4. Whether to retain or eliminate such exams has become a pressing question for many colleges in recent years. For example, the University of California officially eliminated standardized testing requirements in 2020, while other universities like Stanford, Harvard, MIT, and Yale have reinstated them after COVID. While predictive of students' college and post-college performance,[5] standardized exams also reflect—and may exacerbate—preexisting disparities tied to socioeconomic background and other demographic factors.[6] Borghesan (2023), Barron et al. (2024) and Felegi (2024) document effects on eliminating or reducing emphasis of standardized exam scores in college admissions.

universities. Public universities, which typically rank higher, receive substantial government subsidies, making them more affordable but limiting enrollment capacity. Admission to public universities is based equally on a standardized national entrance exam and high school grades. In contrast, private universities, which generally have a higher intake of students, do not require an entrance exam but instead verify proof of high school graduation, admitting students until class capacity is reached. Private universities tend to be more expensive.[7] Based on labor and census data, approximately 25% of the Costa Rican labor force holds a college degree.

The top five universities in the country are:[8] the University of Costa Rica (UCR), the Institute of Technology of Costa Rica (TEC), the National University of Costa Rica (UNA), the Latin American University of Science and Technology (ULACIT) and the Autonomous University of Central America (UACA). Founded in 1940, UCR offers degrees in all areas of knowledge[9] and has an active enrollment of 44,410 students. TEC, founded in 1971, specializes in degrees in agriculture, business and economics, education, and engineering, with 10,552 students enrolled. UNA, established in 1973, offers degrees in all areas except health sciences and enrolls 20,108 students. ULACIT, founded in 1987, provides degrees in all areas except agriculture and basic sciences, health area covers only dentistry, and has an active enrollment of 3,067 students. UACA, established in 1976, offers degrees in all areas except basic sciences, with 2,671 students enrolled.

The first three institutions—UCR, TEC, UNA—are public universities, while ULACIT and UACA are private. Table 1 presents the accumulated number of graduates from these universities between 2000 and 2020, disaggregated by the student's highest degree attained.

Figure 1 shows the average monthly wage (in USD) of all graduates by university over time. The latter highlights that the individuals studied in this paper differ systematically from the general population, not only in terms of education level and academic ability but potentially along other dimensions as well. This distinction raises an important question: are the results presented in this paper specific to the college setting? While the college population is distinct,

7. Annual tuition at private universities averages between $3,000 and $4,000, whereas public universities charge around $1,000.

8. Ranking based on the QS Latin America University Rankings 2023 (https://www.topuniversities.com/university-rankings/latin-american-university-rankings/2023) and several rankings averages (https://www.universityguru.com/universities–costa-rica).

9. Areas of knowledge are grouped into nine categories: (1) Agriculture and Natural Sciences, (2) Basic Sciences, (3) Economic Sciences, (4) Education, (5) Engineering, (6) Health Sciences, (7) Law, (8) Letters and Arts, and (9) Social Sciences.

it exhibits sufficient heterogeneity to allow for subgroup analysis that may be more representative of the general population. For instance, analyzing the effects on graduates from low-income families can provide insights that are closer to broader societal patterns. However, it is important to note that other dimensions, such as age and maturity differences between college and school settings, may make the results specific to higher education.

Table 1: Number of graduates by university and highest academic degree: 2000-2020

Rank	University	Type	Technician	Bachelor	Licentiate	Master	Doctorate	Total
1	UCR	Public	781	27,829	26,801	10,029	4,264	69,704
2	TEC	Public	362	6,615	8,330	3,894	40	19,241
3	UNA	Public	2,014	20,171	11,639	3,896	95	37,815
4	**ULACIT**	Private	456	2,014	3,976	1,422	102	7,970
5	UACA	Private	0	3,406	3,036	239	260	6,941
		Total	3,613	60,035	53,782	19,480	4,761	141,671

Note: This table presents the number of graduates from the top five universities for the period 2000-2020, categorized by their highest academic degree. The "Technician" degree is typically completed in one to two years and is below the bachelor's level. The "Licentiate" degree follows the bachelor's degree and typically requires an additional year of study. Each individual is counted in only one category, based on their highest achieved degree. In instances where a student holds degrees from multiple universities, the university from which the first degree was obtained is assigned for the purposes of this table.

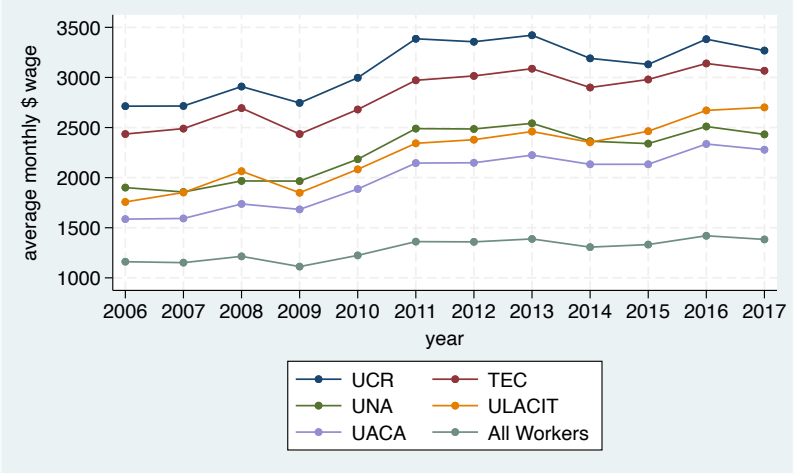


Figure 1: Average monthly $ wage of graduates by university

Note: This figure displays the average monthly wage (in USD) of all graduates from the top five universities in Costa Rica, by calendar year. The data is sourced from labor administrative records.

3. Data

Graduation Records from Top Universities

This dataset consists of historical graduation records from the top five universities in Costa Rica: the University of Costa Rica (UCR), the Costa Rica Institute of Technology (TEC), the National University of Costa Rica (UNA), the Latin American University of Science and Technology (ULACIT), and the Autonomous University of Central America (UACA). For each graduate, this dataset includes a unique identifier to merge with labor data, all degrees earned, graduation dates, program level (technician, bachelor, licentiate, master, doctorate), program name, and field of study / academic area. For four of the universities (UCR, TEC, UNA, and UACA), the admission year is estimated based on the earliest graduation date and the official/historical program duration. For UNA, ULACIT, and UACA, the dataset includes graduation distinctions such as honors. ULACIT provides further information, such as final graduation GPA, exact program start dates, type of high school (public vs private), and high school exit GPA[10] (averaged across core high school subjects in the final two years). A limitation of this study is that it does not capture students who dropped out of college. As a result, we cannot analyze completion rates, and our labor market findings apply only to students who graduated.

National High School Exit Scores

The Ministry of Education of Costa Rica provided an anonymized student-level dataset of high school exit scores for the years 2000 to 2019. This dataset includes the high school math exit GPA (roughly half coming from a standardized exam and half from the last two years in high school), the individual's gender, the type of high school (private or public), and the university the student subsequently attended; the university variable is categorized into six groups corresponding to the top five universities (UCR, TEC, UNA, ULACIT and UACA), along with a residual 'Other' category which includes all other universities and students who did not pursue higher education. Due to anonymization, it is not possible to directly link individuals in this dataset to other ones we use. Instead, we conduct the analysis and estimations

10. Only available for ULACIT students who applied for scholarships, regardless of outcome (51% of ULACIT's historical graduates). This subsample may not fully represent all ULACIT students, but it is valuable for heterogeneity analyses.

within this dataset and subsequently aggregate the data at the university and admission cohort levels to facilitate linkage with other datasets.

Employer-Employee Administrative Dataset

We use a panel dataset collected by the Social Security Administration of Costa Rica, and accessed through the University of Costa Rica, which covers all formal workers in the labor market from January 2006 to December 2017. For each worker, the dataset includes demographic characteristics such as date of birth, age, and gender, as well as monthly labor earnings (wages), three-digit occupation codes, college degree dummy,[11] and additional worker and employer characteristics. Labor earnings are uncensored and are deflated to 2013 real colones using the consumer price index (CPI). The analysis is restricted to full-time workers aged 18 to 65. This dataset does not track informal employment, which represents 39.1% of total employment in Costa Rica, a figure below the Latin American average (ILO 2018). This is less of a concern in our study, as university graduates are much less likely to work informally. According to Costa Rica's National Institute of Statistics (INEC 2015), about 87% of informal workers had at most a high school education, while only 13% had any college education (with or without a degree).

Family Network

This dataset, collected by the National Civil Registry, records the ID of all citizens at birth along with the IDs of their respective parents. To address gaps in older records, which occasionally lack information on one parent, we supplement the dataset with data from the National Matrimony Registry, which documents the IDs of spouses in all recorded unions. By linking this enriched family dataset with the previous labor records,[12] and college records, we are able to assess whether at least one parent holds a college degree for almost half of all historical graduates from the top universities.

11. This proxy variable is generated using the three-digit occupation codes, which in this setting map one-to-one jobs that require college degrees.

12. Recent studies in the area of trade and labor economics have used these two Costa Rican datasets on family networks and labor records. See Alfaro-Ureña et al. (2022), Alfaro-Ureña et al. (2023), Alvarez et al. (2023) and Mendez and VanPatten (2024).

4. Exams to Projects Policy

In 2008, ULACIT University implemented an exogenous change to its academic evaluation policy, transitioning from exam-based assessments[13] to group-project evaluations.[14] This policy, still in effect today, is officially known as *Somos* 100% *proyectos*, 0% *exámenes*.[15] The change was enacted without altering program durations, class calendars, or academic periods. It was approved with minimal discussion or prior notice to the university community, given the private ownership structure of the institution at the time, and implemented in 2008 across all programs, courses, and student cohorts.

The university's owner and Board President, Dr. Silvia Castro Montero,[16] introduced this policy to create a better learning experience more closely aligned with the labor market. The shift emphasized formative (feedback-based) over summative (exam-based) assessments, drawing parallels with the pedagogical principles of "Nueva Escuela." However, unlike ULACIT's approach, "Nueva Escuela" still incorporates quizzes and exams as supplementary tools (Dang et al. (2022)). Table 2 shows how the policy replaced exam-based assessments with project-based evaluations. While some projects may carry the same weight as traditional exams, they still represent a lower-stakes assessment, as they are collaborative, iterative, and accompanied by feedback, allowing students to refine their work over time rather than within a rigid, high-pressure timeframe.

Currently, ULACIT is ranked as the top private university in Costa Rica and fourth overall. No other universities in Costa Rica have adopted a similar policy. The other universities close in ranking presented in Table 1 act as the most natural counterfactual, once we account for any selection bias between universities. A robustness check using only private universities will supplement our analysis. Furthermore, we will account for the possibility of spillover effects onto control university graduates.

13. Based on a pre-2008 sample of syllabi from various academic disciplines, 50% of college grades at ULACIT were determined by exams (quizzes, midterms, and finals), providing context for the magnitude of this shock.

14. A project in this context refers to any form of evaluation that is not an in-person, timed, oral or written assessment (e.g., quizzes, midterms, or finals). These projects often involve long case studies or problem sets, typically collaborative but occasionally individual, and frequently require both written submissions and oral presentations.

15. The evaluation reform was approved by the National Council of Private Higher Education (Conesup) through letter 256-08.

16. Dr. Castro Montero holds a Doctorate in Higher Education Administration from the University of Pennsylvania and a Master's in Education from Harvard University.

Table 2: Course Evaluation in the Business Major: Before and After Policy

Introduction to Statistics (pre-2008)		Introduction to Statistics (post-2008)	
Midterm Exam I	20%	Class Participation	15%
Midterm Exam II	20%	Applied Case Study	45%
Final Exam	40%	Presentation Case Study	15%
Quizzes	20%	Presentation Statistical Index	20%
		Course Survey	5%
Introduction to Accounting (pre-2008)		**Introduction to Accounting (post-2008)**	
Midterm Exam I	20%	Class Participation	15%
Midterm Exam II	20%	Regular Topic Summary Submission	15%
Final Exam	30%	Online Peer Panel Discussion	15%
Homework	30%	Financial Statements Project	40%
		Video Project Results to 'Client'	10%
		Course Survey	5%
Macroeconomics (pre-2008)		**Macroeconomics (post-2008)**	
Midterm Exam I	25%	Class Participation	15%
Midterm Exam II	25%	Analysis Economic Cycle	10%
Midterm Exam III	25%	Research on Measuring GDP	10%
Quizzes/Homework	25%	Analysis Inflation and Unemployment	10%
		Proposal on Macro Policies	40%
		Video on Proposed Policies	10%
		Course Survey	5%

Note: This table presents the course evaluation rubrics for three business major courses, both before and after the exam-to-projects policy change in 2008. The rubrics detail the distribution of grades between various course components, such as exams, projects, and participation. These rubrics were directly extracted and translated from the historical course archive maintained by the National Council of Private Higher Education (Conesup). While the evaluation names post-policy rollout are presented in a simplified format, each evaluation includes several pages of detailed objectives and rubrics that students are expected to follow.

Given the uniqueness of this policy, it is important to assess whether concurrent changes at ULACIT could confound our results. To investigate this, we combined ULACIT's public historical timeline with any policy changes recorded in the comprehensive archives of the National Council of Private Higher Education (Conesup).[17] Table 3 summarizes ULACIT's major policies and events over time. All pre-2008 changes occurred far enough in the past to assume that any potential effects were stable by 2008 and can be captured as part of the university selection bias, which our method will account for; if such assumption were not true that would undermine one of the method's key requirements but we would be able to detect signs of that non-stability in the results' plots as we will explain in a later section. Furthermore, the 2013 Technology Center benefited specific programs and so we will be able to show that the main results are not driven by these majors.

17. Private universities in Costa Rica must submit detailed proposals to Conesup for approval before implementing major policy changes, such as modifications to programs, curricula, or evaluation methods. These proposals document both current and previous policies.

Table 3: ULACIT General Timeline

Year	Event and Details
1987	Foundation of ULACIT by Dr. Álvaro Castro Harrigan.
1999	Mandatory sequence of four English courses in all programs.
2000	Dr. Silvia Castro becomes President of ULACIT University.
2001	Updated admission rules and procedures.
2003	English language test made a graduation requirement.
2008	Elimination of exams for group projects.
2013	Opening of the Innovation and Technology Transfer Center (CIT).

Note: This table presents a timeline of key events in the history of ULACIT. It combines information from ULACIT's public historical timeline with policy changes documented in the National Council of Private Higher Education (Conesup) archives.

Figure 2 presents some graduation statistics before and after the policy roll-out. Given that the policy applied to all classes simultaneously in 2008, exposure to treatment is defined at the graduation level. Graduates prior to 2008 were entirely unexposed, having taken all their courses under the traditional evaluation system. By contrast, Figure 2a illustrates how the exposure of graduating cohorts to the policy increased over time. By 2013, most graduates were fully exposed to the new evaluation regime, defined as students who began their studies in 2008 and completed all coursework under the revised policy.

Figure 2b reveals a decline in average program duration following the policy roll-out. The shift to project-based evaluations allowed students to progress more efficiently through coursework, as project timelines are typically more flexible than exam schedules. Importantly, the standard undergraduate program's official duration (approximately three years) and course structure remained unchanged during this period. By comparison, Conare (2017) and Conare (2023) report that average undergraduate program duration at UCR-TEC-UNA had remained steady at 4.5 to 4.9 years.

As shown in Figure 2c, average grades increased post-policy. This trend may reflect several factors: increased student control over grades via project-based evaluations (project grade inflation),[18] cross-country grade inflation,[19] or enhanced learning outcomes. Higher grades, and

18. Bowden et al. (2023) discuss potential unintended consequences of policies encouraging academic leniency.

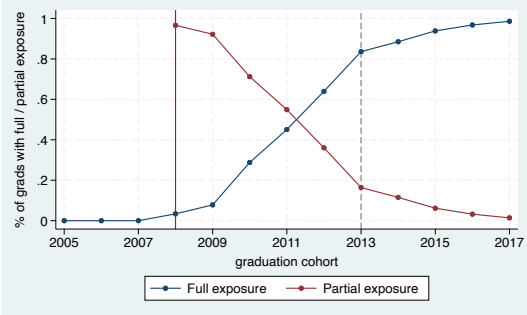
19. Rampell (2011) document widespread grade inflation in the US, especially at private and selective universities. However, in this context, GPA increases align closely with policy exposure. Moreover, in this setting, honor distinctions are a fixed function of minimum threshold grades, and Figure 8 in the Appendix shows that there is no unexpected spike on this variable for other untreated universities that coincides with the treatment date; this provides indirect evidence that cross-country grade inflation unrelated to projects is not likely what drives the higher grades in ULACIT after policy roll-out.

hence lower failing rates, could also explain shortening program durations; however, given that the grade distribution average is well over the passing grade of 70, this is unlikely the primary mechanism. Furthermore, the policy appears to have compressed the grade distribution, as evidenced by a decrease in GPA variance. Figure 2d shows a spike in the number of honor distinctions after 2008 as a reaction to the shift in the grade distribution.

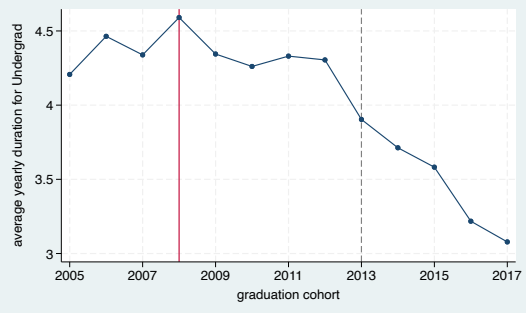
Even though treatment exposure is defined based on graduation timing in this context, it is inherently endogenous: students have discretion over when to complete their programs, and the reform itself may have altered graduation dynamics at ULACIT—for example by shortening program duration or, conversely, by postponing graduation. To address this, we instead map treatment exposure to students' admission year, which provides a more exogenous timing variable. One caveat is that the admission years of the control universities are estimated from graduation dates, which could transmit some of the same endogeneity if graduation dynamics in those institutions were shifting. Given earlier evidence that program durations at public control universities were remarkably stable over this period, the graduation-to-admission mapping is unlikely to generate systematic bias; under classical measurement error, any residual misclassification would attenuate rather than create treatment effects. Importantly, this concern does not apply to ULACIT—the treated university—where we directly observe admission years from administrative records. Since graduation dynamics did change at ULACIT, having exact admission data there ensures that treatment assignment is measured without error precisely where it matters most.

Figure 2e illustrates the admission year of the first cohort with at least one year of exposure to the new evaluation system. Due to individual heterogeneity in progression rates, graduating students in the same year often have different admission years. The figure indicates that the mode admission year was 2005, which we designate as the new event year.[20] Shifting the analysis from graduation to admission cohort introduces some fuzziness in the treatment definition but mitigates the endogeneity associated with graduation timing. Consequently, this approach yields an Intention-to-Treat (ITT) estimate that accounts for variations in treatment exposure across admission cohorts.

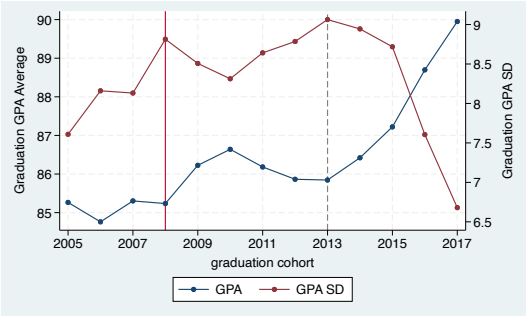
20. Table 3 shows a new English graduation requirement starting in 2003, however, this does not conflict with our event of interest as the former is at the level of graduation cohorts, not admission; all ULACIT admission cohorts we include in this paper's estimations have passed this requirement.



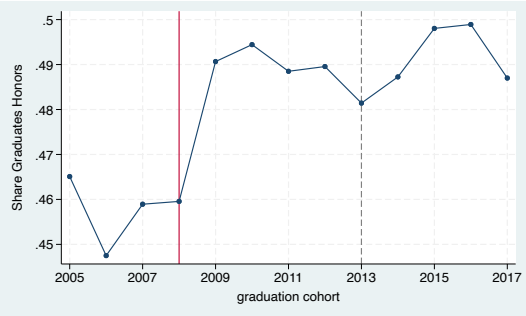
(a) Share of graduates by type of exposure



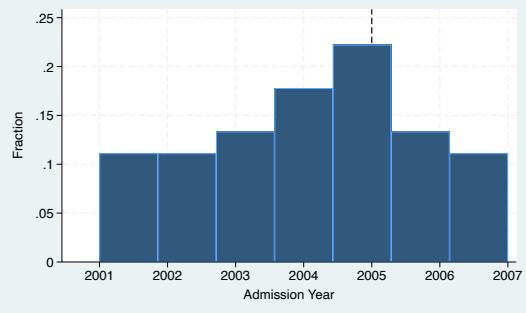
(b) Average Duration in Program



(c) Average Graduation GPA and SD



(d) Share Graduates with Honors



(e) First cohort with one year of exposure

Figure 2: ULACIT Policy Statistics

Note: This figure presents time series data for ULACIT university graduating cohorts, broken down as follows: (a) the share of graduates who completed all their courses under the new no-exams policy versus those with partial exposure; (b) the average duration (in years) of graduates' time spent in the undergraduate academic program; (c) the average graduation GPA along with the standard deviation; (d) the share of graduates who earned a graduation honor distinction; (e) the admission year of the first cohort to experience at least one year of exposure to the new evaluation system. The solid vertical line in 2008 marks the introduction of the new evaluation system without exams, and the corresponding dotted line in the first four panels indicates the point when most graduating students at ULACIT were fully exposed to the system.

5. Conceptual Framework

The elimination of exams and the shift to project-based evaluations can influence outcomes through several distinct but interrelated pathways. For clarity, we group these into three pri-

mary domains: student composition, faculty responses, and graduates' labor market outcomes. We then interpret the observed labor market effects through a set of complementary, though not exhaustive, mechanisms, treating all of these domains as policy-induced responses rather than independent confounds.

i. Student Composition. The policy may have influenced who chose to enroll at ULACIT. Prospective students could have responded to the removal of exams based on perceived rigor, pedagogical preference, or career alignment. We study how this affected observable characteristics such as academic ability (proxied by high school math exit GPA), gender, degree level (undergraduate vs. graduate), and socioeconomic background (proxied by high school type and parental education). Because the policy was enacted after many students were already enrolled, we can cleanly separate policy effects from compositional shifts by focusing on pre-2008 admission cohorts. In a later section, we provide indirect evidence that dropout rates did not rise among the partially treated cohorts, minimizing concerns about survivor bias and permitting this strategy.

ii. Faculty Responses. The elimination of exams required instructors to redesign assessments and provide more continuous feedback, increasing workload. This pedagogical shift may have affected faculty satisfaction and retention, particularly among instructors opposed to the reform or overwhelmed by its demands.[21] We examine changes in faculty turnover and wages to assess this response.

iii. Labor Market Outcomes. We evaluate several labor market outcomes—including formal employment and wages—and interpret the observed wage effects through several possible mechanisms:

- *Work Experience during College*. The project-based system may better align with real-world work environments, indirectly encouraging students to begin working earlier, often through internships that transition into full-time jobs.

- *Soft Skills*. Group-based evaluations may strengthen skills such as teamwork, communication, and leadership. While difficult to measure directly, we proxy these effects using alumni networks—defined as the number of college peers working in the same

21. Ainsworth et al. (2024) shows how structural changes in school settings, such as a four-day week, can increase teacher turnover.

firm[22]—and managerial job classifications.

- *Signaling and Grade Compression*. The removal of exams and subsequent grade inflation may reduce the informational value of transcripts and honors distinctions. For example, the traditional honor wage premium[23] could be weakened—though ULACIT already exhibited little pre-policy honor premium, likely due to the high share of honors distinctions (Figure 2d). The compression of grades may also help lower-performing graduates gain access to better-paying firms. We explore this channel using firm fixed effects.

- *Occupational Re-sorting*. Graduates may transition into occupations that better leverage collaborative and problem-solving skills developed under the new evaluation regime. We test this using three-digit occupational codes fixed effects.

- *Learning*. Although direct measurement of learning is unavailable, some portion of the wage gains may reflect real academic improvements. In the absence of standardized post-college assessments in Costa Rica,[24] we interpret unexplained residual wage effects as potential indirect evidence of this scenario. Prior studies suggest that project-based models can improve learning outcomes (Saavedra et al. 2022; Krajcik et al. 2023).

6. Empirical Strategy: Differences-in-Differences

In this section, we outline our approach to estimating the impact of the policy of interest. Specifically, we employ a differences-in-differences (DiD) strategy, comparing admission cohorts from ULACIT to those of other top universities, before and after the policy rollout. Our identification relies on three main assumptions: (1) parallel trends, (2) no anticipation, and (3) no simultaneous shocks coinciding with the event of interest.

The first assumption, parallel trends, will be assessed and discussed in detail when presenting the estimation results in the next section. Because this assessment relies on the event-study plots and on the clustering strategy developed below, we defer its evaluation until then. This assumption ensures that pre-event differences across universities remain statistically constant

22. This follows the literature on social capital and networks, e.g., Zarate (2023).
23. See Freier et al. (2015), Feng and Graetz (2017), Pietro (2017), Khoo and Ost (2018), and Busso et al. (2023).
24. Alternatives such as the Law Bar Exam, Accounting Bar Exam, or Medical Board Certification cannot be used in this setting due to post-2008 implementation, missing scores, near-zero failure rates, or because ULACIT does not offer the corresponding major.

over time, allowing us to isolate and adjust for selection bias in post-event differences. The second assumption is justified by the exogenous nature of the policy, which was rapidly approved and implemented with minimal prior notice to the university community—enabled by the institution's private ownership structure at the time. This feature allows us to disentangle compositional changes from the policy's core effects by focusing on pre-2008 admission cohorts. The third assumption is supported by the timeline presented in Table 3 and its accompanying discussion.

Recent DiD literature emphasizes the need to account for time-heterogeneous treatment effects in staggered event settings (Sun and Abraham (2021)). However, since this policy was implemented in a single rollout, we adhere to the standard DiD model (Roth et al. (2023)). Our unit of analysis in the dynamic DiD framework is admission cohorts rather than calendar years. Consequently, this approach inherently accounts for heterogeneity in treatment effects, as exposure to the policy varies by cohort.

Model Specifications

Compositional Changes

We estimate compositional changes, formal sector employment probability, and likelihood of working before graduation using the following model:

$$Y_i = \alpha + \sum_{\substack{k=2001 \\ k \neq 2007}}^{2012} \theta_k D_i^k + \delta X_i + \epsilon_i \tag{1}$$

Here, Y_i represents graduate i's outcome, which includes variables such as enrollment type (undergraduate or graduate), high school math exit GPA, a male dummy, academic area dummies, a private high school dummy, a parental college dummy, a pre-graduation work dummy, and a formal sector job dummy. D_i^k are admission event dummies that activate for ULACIT graduates enrolled in year k. The coefficient for the 2007 admission dummy ($k = 2007$) is normalized to zero since 2008 marks the policy event year for compositional margins. Formal employment and pre-graduation work exposure outcomes use the 2005 event year as these outcomes can react even for partially exposed students. Admission cohorts beyond 2012 are excluded from the estimations, as the latter is the most recent cohort to have graduated and en-

tered the labor market by the final year covered in our labor records. X_i includes time-invariant controls such as admission year and university fixed effects. The coefficients of interest, θ_k, capture the policy's impact on ULACIT graduates as deviations from the pre-event year gap between treated and untreated units.

Soft Skills and Wages

The model for soft skills and wages is specified as:

$$Y_{it} = \alpha + \sum_{\substack{k=2001 \\ k \neq 2004}}^{2012} \theta_k D_i^k + \delta X_i + \epsilon_{it} \tag{2}$$

In this case, $connect_{it}$ represents the number of college classmates from graduate i's cohort who are working at the same firm as them in year-month t. In other words, it captures how many of a graduate's current co-workers attended college with them. $manager_{it}$ indicates if graduate i's job at time t is classified as managerial, and $wage_{it}$ represents the log monthly wage of graduate i.[25] In our main specification, we pool all available post-graduation monthly outcome observations per individual. This approach increases statistical power and allows us to estimate the average treatment effect across the full observed post-graduation window. Similar to the previous model, D_i^k are admission event dummies, with $k = 2004$ normalized to zero since the policy was implemented in 2005. X_i includes time-invariant controls such as admission year and university fixed effects. For certain outcomes—such as teacher wages and turnover—this model adjusts the x-axis to calendar years, unit i to faculty instead of graduates, and sets the event year as 2008. We adopt the latter, as we cannot link individual instructors to specific students and their admission cohorts; accordingly, calendar year is the more appropriate time dimension. Turnover is measured with a dummy that activates on the last year a teacher works for the university.

Static DiD Wage Model

The static DiD wage model follows this equation:

$$wage_{it} = \alpha + \theta D_i + \delta X_i + \epsilon_{it} \tag{3}$$

25. We exclude all jobs held by graduates before their college graduation date from these regressions. However, indicators of early work exposure were constructed prior to omitting the corresponding wage data.

Here, D_i activates for ULACIT graduates admitted after 2004. We will conduct heterogeneity analyses by gender, academic level, high school type,[26] high school exit GPA,[27] and parental education.[28]

Error Clustering

The ideal approach would be clustering at the treatment level (i.e., university), accounting for within-university correlations over time. However, with only five universities, this approach risks biased standard errors and over-rejection of the null hypothesis. While robust inference methods like wild cluster bootstrapping can address the small number of clusters issue, they may still produce noisy estimates. Alternatively, clustering at the student level avoids the previous few clusters issue but neglects correlations among students within the same university. Clustering at the university-cohort level balances these concerns by accounting for within-cohort correlations and increasing the number of clusters (approximately 60, based on five universities and 12 cohorts). Since this level aligns with varying cohort exposures, we cluster errors at the university-cohort level for all outcomes and provide sensitivity checks with student-level clustering where applicable (e.g., wages). Given that only one university received treatment, we proceed with caution in interpreting statistical inference.

7. Results

Our identification strategy relies on the parallel-trends assumption. While this assumption cannot be tested directly, we probe it by examining the pre-treatment event-study coefficients. Pre-event coefficients that are small and statistically indistinguishable from zero suggest stable pre-period gaps and provide indirect support for parallel trends. As Roth (2022) emphasizes, formal joint tests of pre-trends have low power and are difficult to interpret, recommending instead that researchers draw on context-specific economic knowledge when discussing potential violations.

26. This variable is available only for ULACIT graduates. For this analysis, we filter ULACIT graduates by this variable while keeping the control group fixed.

27. This variable is available only for ULACIT graduates who applied for scholarships or financial assistance, regardless of outcome (historically, 51% of all graduates). While not fully representative, it allows us to identify top high school students. For this analysis, we filter ULACIT graduates by this variable while keeping the control group fixed.

28. This variable is available for almost half of all historical graduates from the top universities.

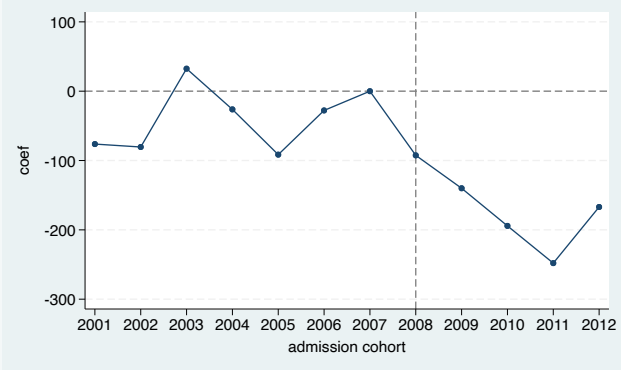
Rather than relying on the significance of pre-tests, the Rambachan and Roth (2023) HonestDiD framework evaluates whether post-treatment violations of parallel trends are not too large relative to pre-treatment deviations. Although this offers a more comprehensive approach, the clustering strategy in our setting raises concerns about standard errors, limiting our ability to rely heavily on these tests. Nevertheless, for transparency, we will report the main DiD estimates using the HonestDiD procedure in the Appendix (Figures 12 and 13), which display robust confidence intervals for the bundled post-treatment effect under increasing levels of potential assumption violations.

Given these limitations, we assess the assumption primarily by examining the magnitude of the estimates, their overall stability before the event, and the extent of changes thereafter, while also checking whether individual pre-treatment coefficients are statistically insignificant. As shown in the plots that follow, these conditions are generally satisfied across specifications.

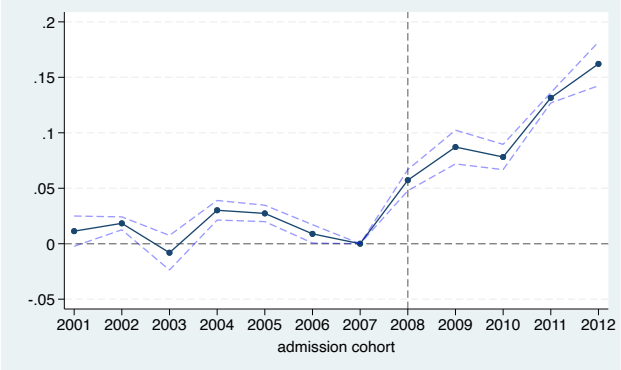
Student Composition

We begin by examining the impact of the exam-to-project substitution on ULACIT's student composition using Figure 3. Figure 3a shows a 4.2% decline in total enrollment at ULACIT relative to the control group, proxied here by the number of graduates given the records available for this project. No confidence interval is shown, as the figure reflects a descriptive statistic rather than an estimate. This decline in enrollment was likely driven by hesitation or deterrence among prospective students facing an unfamiliar and unconventional evaluation system. It is unlikely to reflect increased dropout rates, which would have appeared earlier, among the partially treated cohorts.

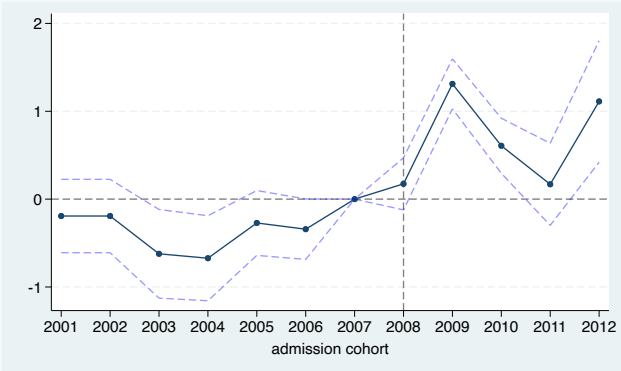
Because we do not directly observe dropout, we rely on indirect evidence: the number of graduates between 2005 and 2008 remained stable, suggesting that no meaningful rise in dropouts occurred and that survivor bias among the partially treated cohorts is minimal. The broad stability of compositional variables prior to the 2008 admission cohort we will present and discuss further strengthens this interpretation. Later on, this will allow the use of pre-2008 cohorts to disentangle compositional changes from other channels.



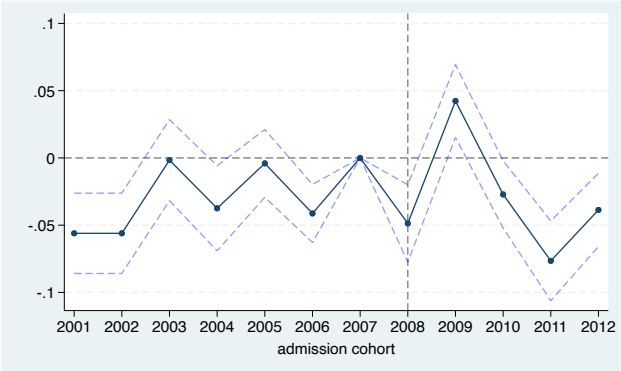
(a) Enrollment Proxy: Number of Graduates



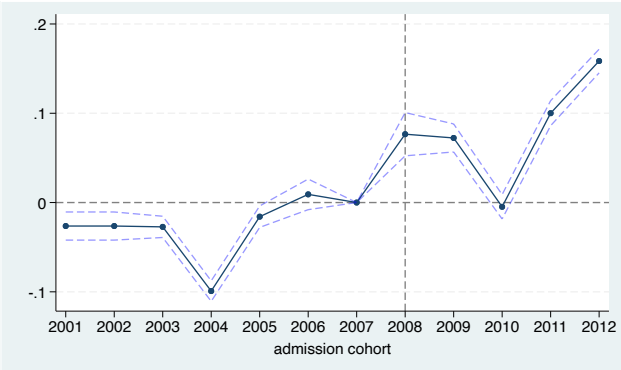
(b) Share Enrollment Grad Program



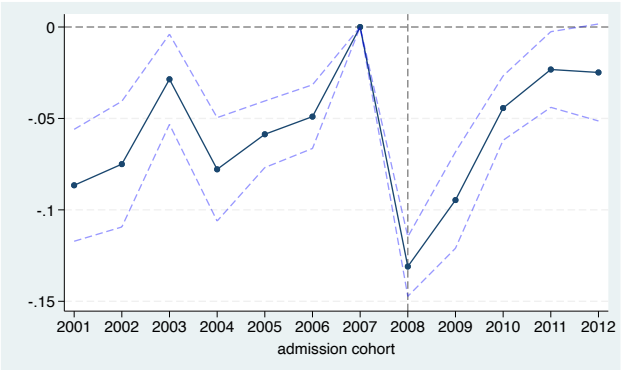
(c) HS math exit-GPA



(d) Share of Male



(e) Share of Private HS



(f) Share of Parents College

Figure 3: Admission Cohort Composition DiD

Note: This figure presents difference-in-differences estimates of student composition at ULACIT vs control universities by admission year. Panel (a) shows total number of graduates, which we will take as a proxy for enrollment; because this is a descriptive statistic rather than an estimation, no confidence intervals are shown for (a); panel (b) displays graduate school enrollment as a share of total enrollment; panel (c) shows the average math proficiency of incoming cohorts, measured by their High School math exit GPA; panel (d) shows the share of male students; panel (e) displays the share of admitted students from private high schools; panel (f) shows the share of admitted students with college-educated parents. All estimates are normalized to the 2007 gap. The dotted line marks the first cohort of students who enrolled at ULACIT with full knowledge of the exam-to-project policy rollout.

Figure 3b shows a 9.3 percentage point increase in graduate enrollment as a share of total enrollment. This suggests that graduate programs, which were already less reliant on exams, were better aligned with the policy change. This relative increase reflects not only a decline in undergraduate enrollment but also a rise in graduate enrollment from new students outside ULACIT, possibly perceiving the institution as consolidating its reputation as a graduate-focused university.

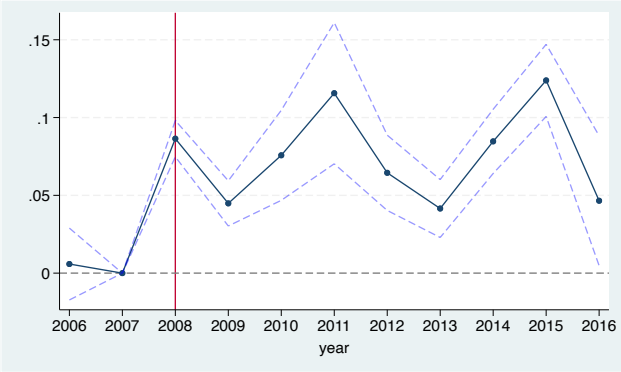
The institution's primary concern was the potential attraction of lower-performing students. However, Figure 3c shows that math proficiency among incoming cohorts increased by one point (out of 100) following the policy change. At the same time, Figure 3d indicates no significant shift in gender composition. Figure 3e further suggests a 10 percentage point rise in the share of students from private schools, which usually serve the top income households, alongside tentative—albeit noisy—evidence of a temporary decline in enrollment of students with college-educated parents, who may have advised caution, as shown in Figure 3f.

Faculty Responses

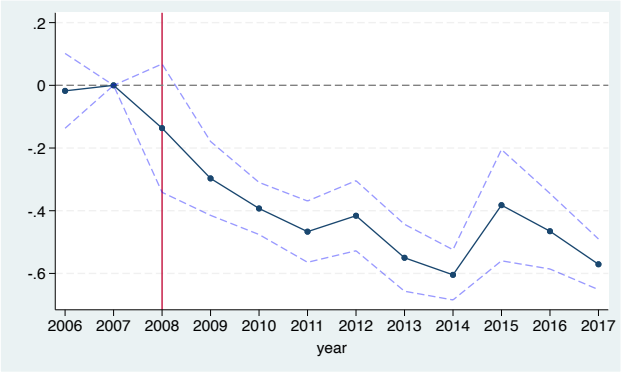
Next, we examine faculty responses to the policy using Figure 4. For these outcomes only, we use calendar year as a more appropriate time dimension because we cannot link instructors to specific students and their admission cohorts. This also implies a shorter pre-period than before as our labor data begins in 2006. Figure 4a shows a marked increase in teacher turnover, with a 7.3 percentage point rise. Anecdotal evidence suggests that this increase was driven by faculty disagreements with the new system's vision and concerns over the heavier workload associated with project-based assessments. Unlike traditional exams, which allow for standardized grading, project-based evaluations require more qualitative assessment and individualized feedback, significantly adding to faculty responsibilities. This shift likely made teaching more time-intensive, prompting some professors to leave.

To understand who exited, we compare observable characteristics of leavers versus stayers. We find no difference in average age, but leavers were 9 percentage points less likely to hold full-time positions and 9 percentage points less likely to be ULACIT alumni. They were also 4 percentage points less likely to hold a graduate degree—potentially reflecting less experience with research and project-based work. These patterns suggest the policy did not disproportionately drive out the most experienced or highly credentialed faculty, but rather

those with weaker institutional ties and those more exposed to the policy's demands. While higher turnover introduces potential instability, it may also have facilitated a transition toward instructors better aligned with project-based pedagogy, fostering a more adaptive teaching environment.



(a) Share of Faculty Exit



(b) Faculty Wage

(c) # Faculty

Figure 4: Faculty Response

Note: This figure presents difference-in-differences estimates for faculty at ULACIT vs control universities by calendar year. Panel (a) shows the share of faculty to exit employment at the university each year; panel (b) displays the average faculty wage; panel (c) shows the total number of faculty; because this is a descriptive statistic rather than an estimation, no confidence intervals are shown for (c). All estimates are normalized to the 2007 gap. The solid vertical line in 2008 marks the introduction of the new evaluation system without exams.

Faculty wages at ULACIT stagnated following the policy change, largely due to slowed student enrollment and resulting fiscal constraints. Over time, this persistent stagnation led to large salary disparities across universities, as illustrated in Figure 4b. The widening pay gap likely made it more difficult for ULACIT to attract and retain high-quality faculty. At the same time, increased workload, faculty objections to the new academic vision, and wage stagnation constrained faculty growth. Faculty size remained stagnant, a trend that compounded in relative terms over time, as seen in Figure 4c; no confidence interval is shown in this plot, as it

represents a descriptive statistic rather than an estimation. By 2017, a decade after the policy change, ULACIT had missed the opportunity to further increase its faculty size, potentially limiting its capacity to expand course offerings and maintain institutional competitiveness.

Soft Skills

Turning now to soft skills, Figure 5 highlights a relative increase for ULACIT graduates in the number of co-workers who attended college with them (alumni networks). Different definitions of college classmates are considered: Figure 5a uses the same admission year and broad academic area, while Figure 5b refines this to the same admission year and specific major. Figure 5c normalizes the latter connections by cohort size, revealing a 6.5 percentage point increase, which may be related to the enhanced social interaction fostered by the group project evaluation system.[29]

However, this increase in network size did not translate into a comparable rise in managerial positions for ULACIT graduates, per Figure 5d, suggesting that while the policy may have fostered stronger peer connections, it did not directly enhance leadership development. Alternatively, leadership effects may be manifesting through other, less direct channels.

29. An alternative explanation is that the policy reduced the informativeness of academic records, raising the cost for employers to evaluate individual candidates. In response, only a subset of firms may choose to invest in interpreting ULACIT's new system, leading to a concentration of graduates within those firms. Within this smaller group, employers may then rely more heavily on known pipelines or alumni referrals, reinforcing peer clustering. This employer-side sorting mechanism is not mutually exclusive with our soft skill interpretation and may in fact operate alongside it.

(a) # co-workers from Area Cohort

(b) # co-workers from Major Cohort

(c) Connections normalized by MajorCohort size

(d) Managerial Job

Figure 5: Soft Skills: Alumni Networks and Leadership

Note: This figure presents Dynamic Difference-in-Differences estimates of the effect of the exam-to-project policy on alumni networks and managerial outcomes, with the 2004 gap normalized to zero. Panel (a) shows the estimated impact on the number of co-workers from the same broad academic area and admission cohort, while Panel (b) refines this to co-workers from the same major and admission cohort. Panel (c) normalizes the latter measure by cohort size, capturing relative increases in peer connections. Panel (d) examines the effect on managerial job attainment. The solid vertical line in 2005 marks the introduction of the new evaluation system without exams, and the corresponding dotted line marks the first cohort of students who enrolled at ULACIT with full knowledge of the exam-to-project policy rollout and thus received full exposure to the system.

Labor Outcomes

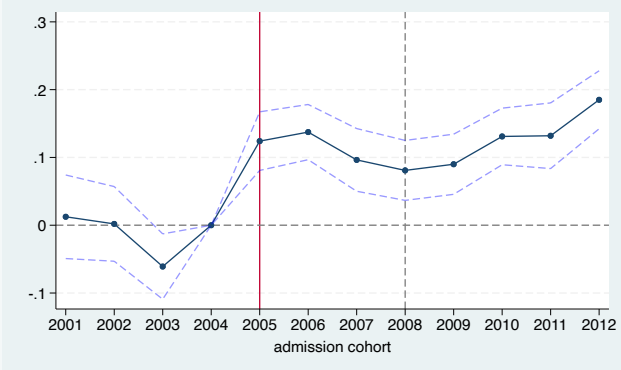
Figure 6 examines the policy's effects on labor market outcomes. Figure 6a shows an 13.8 percentage point increase in the probability of working during college.[30] This increase aligns with better integration between educational methods and workplace dynamics. The new evaluation system also provided greater flexibility, incentivizing students to explore work opportunities earlier. Figure 6b shows an 8.3 percentage point increase in the probability of working in

30. Although the variable is defined as a dummy for working before graduation, the observed jump in 2005 rather than 2008 suggests that the work occurs during college, as pre-college outcomes for the 2005–2007 cohorts are fixed and unaffected by the policy.

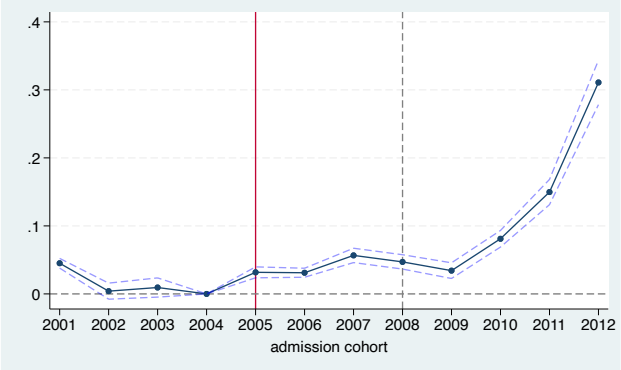
the formal sector, which was a primary objective of the policy. If our dataset included both formal and informal workers, the increase in formal employment could itself explain part of any observed wage gains, since workers typically earn more in the formal sector. However, because we only observe wages for formal workers, this channel cannot be directly captured. Instead, selection into formality works against us: as marginal graduates enter formal jobs, they are likely negatively selected relative to existing formal workers, which biases our estimated wage effects downward. Our wage results should therefore be interpreted as conservative estimates.

Figure 6c shows that the first partially treated cohort—those with at least one year of exposure—already experiences an immediate wage gain. The fact that effects emerge right away, rather than building smoothly from zero, is consistent with mechanisms that operate quickly: for instance, work experience during college and alumni networks can begin to pay off with only limited exposure late in the college program. Supporting this interpretation, LeBarbanchon et al. (2023) provides experimental evidence that even short one-year work experiences during school can generate significant and rapid improvements in students' labor market outcomes. As exposure intensifies across subsequent cohorts, the estimated effects rise monotonically. The 2008 admission cohort, which is the first to be fully treated while still largely unaffected by compositional changes, exhibits an 18% wage increase and serves as our cleanest benchmark for the overall effect of the reform. This large magnitude is consistent with multiple reinforcing mechanisms, including enhanced soft skills and greater work experience during college, as we have seen so far, along with other factors related to access to better-paying firms and occupational re-sorting, which we will discuss shortly.

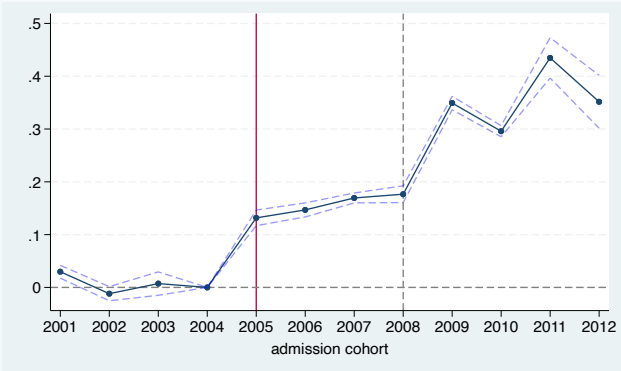
Beginning with the 2009 admission cohort, a further 20 percentage point wage gain emerges, largely driven by compositional shifts toward higher-ability students, those from wealthier backgrounds, and an increased share of graduate-level enrollment. These later gains likely represent an upper bound, as ULACIT's early adopter status allowed its graduates to differentiate themselves on multiple fronts.



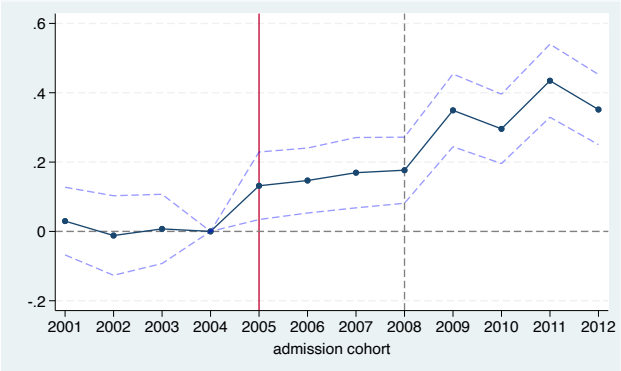
(a) Work before Graduation Dummy



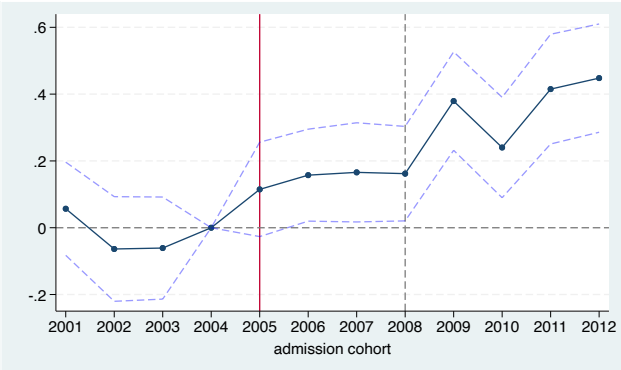
(b) Formal Work Dummy



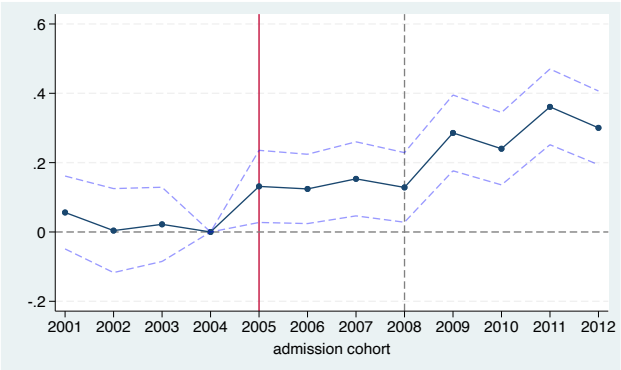
(c) Wage: university × cohort clustering



(d) Wage: student clustering



(e) Wage: Only Private Universities



(f) Wage: Restricted Control - No Spillover

Figure 6: Labor Outcomes DiD

Note: This figure presents Dynamic Difference-in-Differences estimates of the exam-to-project policy's effect on labor market outcomes. Panel (a) shows impacts on working during college; Panel (b) shows effects on formal employment. Panels (c) and (d) report wage effects using university×cohort and student-level clustering, respectively. Panel (e) tests robustness by restricting to private universities. Panel (f) tests the robustness of the wage effects by restricting the control sample to graduates who never worked alongside treated ULACIT graduates—defined as being in the same firm and occupation at the same time—and were therefore not directly exposed to the policy or subject to potential spillover effects. The solid vertical line in 2005 marks the introduction of the new evaluation system without exams, and the corresponding dotted line marks the first cohort of students who enrolled at ULACIT with full knowledge of the exam-to-project policy rollout and thus received full exposure to the system.

Figure 6d confirms the statistical significance of the wage gains using student-level clustered standard errors,[31] while Figure 6e demonstrates the robustness of the effect when restricting the sample to private universities. Figure 9 in the Appendix shows that the effect is not driven by academic majors that benefited from the 2013 Technology Center infrastructure project, and Figure 10 shows a robustness check when controlling for calendar year fixed effects and cumulative (time-varying) labor market experience, measured in years. Additionally, Figure 11 in the Appendix presents a placebo test in which one control university is assigned a pseudo-treated status, ULACIT is excluded, and the remaining three serve as controls. The lack of any systematic wage divergence in this placebo test supports the validity of our control group.

To assess potential bias from spillover effects onto control units, we consider two possible channels. First, spillovers could be positive if control graduates benefit from working alongside ULACIT graduates, for example through the transfer of collaborative or social skills, which would raise their own wages and attenuate the estimated treatment effect. Second, spillovers could be negative if ULACIT graduates crowd out comparable peers in the labor market, generating wage losses for control graduates and thereby inflating the estimated treatment effect, as in Crepon et al. (2013).

To examine the empirical relevance of these mechanisms, Figure 6f re-estimates the wage effect after excluding control graduates who ever worked at the same firm, in the same occupation, and during the same year-month as a treated ULACIT graduate—thus removing those most likely to have been directly exposed to spillovers. The resulting estimates are very similar for the 2005–2007 cohorts and slightly smaller for later cohorts, which is more consistent with negative spillovers onto control units than with positive ones. Based on this, we focus our remaining analysis on the pre-2008 cohorts, who were not subject to self-selection based on the policy and experienced minimal exposure to spillovers.

Table 4 decomposes the policy's effect on wages (excluding compositional effects and clustering errors at the university-cohort level) into potential mechanisms by sequentially introducing intermediate channels into the regression. A key challenge in this decomposition is that some mechanisms may be correlated and may operate through one another, making it difficult to isolate their independent effects. For example, part of the alumni network effect may

31. The tighter confidence intervals under university-cohort clustering may reflect few clusters bias, as discussed in the methods section.

be captured by firm fixed effects, as internal connections can facilitate access to better firms. Similarly, work experience during college may also influence firm fixed effects. To account for these interactions, we introduce mechanisms in a thoughtful sequence: work experience during college, alumni networks, firm fixed effects, and finally, occupational fixed effects. This ordering implies that any wage gains attributed to firm fixed effects reflect access to better firms that is unrelated to work experience during college or alumni networks but rather stems from grade camouflage or other unmeasured soft skills.

Table 4: Policy effect on Wages and Mechanisms: Without compositional changes

	(1) Wage β/(SE)	(2) Wage β/(SE)	(3) Wage β/(SE)	(4) Wage β/(SE)	(5) Wage β/(SE)
θ- effect	.140***	.100***	.072***	.055***	.047***
	(.012)	(.011)	(.008)	(.014)	(.013)
Work bf Grad	NO	**YES**	YES	YES	YES
Networks	NO	NO	**YES**	YES	YES
Firm FE	NO	NO	NO	**YES**	YES
Occupation FE	NO	NO	NO	NO	**YES**
Adj-R^2	0.025	0.033	0.067	0.338	0.396
# Obs (K)	3,408	3,408	3,408	3,408	3,408
# FE	11	11	11	13,433	13,540
# Grads (K)	41	41	41	41	41
Avg $ W	2,206	2,206	2,206	2,206	2,206

*Note: This table shows the estimation of the effect of the exam-to-projects policy on wages in column (1), using pre-2008 admission cohorts only (no selection into policy). The baseline controls include admission year and university fixed effects. Subsequent columns add the following controls: a dummy for working before graduation (during college), number of college classmates from graduate i's major cohort who are working at the same firm as them in year-month t (alumni networks) normalized by the major-cohort size, firm fixed effects, and 3-digit occupation fixed effects. Standard errors are reported in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.*

We estimate that roughly 29% of the wage gain may be associated with greater work experience during college, which appears to help graduates secure a higher initial post-graduation wage and access to better firms. About 20% is linked to alumni networks, which may facilitate better job placements and career advancements within firms. An additional 12% is absorbed by firm fixed effects, plausibly capturing access to better firms through grade compression, signaling, or other acquired soft skills. Occupational re-sorting toward roles better aligned with project-based skills accounts for around 6% of the wage gain. The remaining 33% is unexplained and is consistent with other unmeasured factors, such as enhanced learning or the development of valuable soft skills—including negotiation, communication, and teamwork.

It is worth noting that these decomposition exercises focus on a subset of mechanisms we can measure directly at the individual or student level. Other important channels—such as

the faculty responses like turnover documented earlier—may also mediate the reform's long-run effects, but cannot be cleanly incorporated into this regression-based framework. These decomposition estimates should therefore be viewed as partial, highlighting several key pathways while leaving open the possibility of additional institutional or behavioral mechanisms at play.

Heterogeneous Effects

Table 5 breaks down the wage effects (excluding compositional effects) by various characteristics. The wage gain is significantly higher for male students, driven by the unexplained component of the mechanism, despite their female counterparts experiencing greater gains through alumni connections and work experience during college. The wage gain comes entirely from undergraduate students, as graduate programs were already aligned with the policy, making them a robust placebo test.

Table 5: Policy effect on Wages and Heterogeneity: Without compositional changes

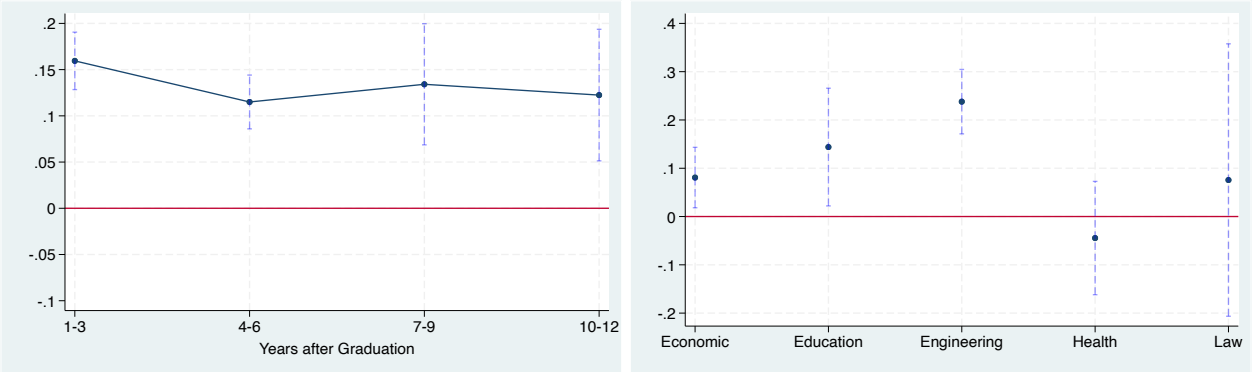
	(1) Wage β/(SE)	(2) M	(3) F	(4) U Enroll	(5) G Enroll	(6) Prv HS	(7) Pub HS	(8) H-GPA HS	(9) L-GPA HS	(10) Uni Prnt	(11) No Uni Prnt
θ- effect	.140***	.188***	.088***	.145***	.025	.072	.150***	.143***	.205***	.176***	.150***
	(.012)	(.016)	(.028)	(.012)	(.121)	(.053)	(.011)	(.031)	(.028)	(.026)	(.047)
Adj-R^2	0.025	0.027	0.021	0.026	0.006	0.024	0.025	0.024	0.023	0.027	0.034
# Obs (K)	3,408	1,573	1,835	3,267	141	3,275	3,345	3,256	3,288	611	727
# FE	11	11	11	11	11	11	11	11	11	11	11
# Grads (K)	41	19	23	39	2	40	40	40	40	9	9
Avg $ W	2,206	2,358	2,075	2,104	4,565	2,206	2,206	2,206	2,206	2,186	1,920

*Note: This table shows the estimation of the effect of the exam-to-projects policy on wages in column (1), using pre-2008 admission cohorts only (no selection into policy). The baseline controls include admission year and university fixed effects. Column (2) and (3) disaggregate the estimation by gender, while column (4) and (5) disaggregate by type of enrollment (undergraduate vs graduate). Column (6) and (7) repeat the baseline estimation, filtering by type of high school (public vs private). Column (8) and (9) filter by high school exit-GPA, above (H) or below (L) the median value for the respective admission year at ULACIT. The variables high school type and exit-GPA are available only for ULACIT graduates; for this analysis, we filter ULACIT graduates by this variable while keeping the control group fixed. Finally, Column (10) and (11) disaggregate by parents' college education. Parental variables are available for almost half of the full sample of graduates, as permitted by the Civil Registry dataset. Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.*

The effect is concentrated among students from public high schools, which predominantly serve households in the bottom four income quintiles. Public high school students experience a greater increase in work exposure during college and improved access to better-paying firms compared to their private high school counterparts, likely because wealthier students have less need to work during their education and already have access to top firms. The effect is also

larger for students below the median high school exit GPA, partially driven by better access to top firms—consistent with the benefits of grade compression and academic camouflage. The effect does not vary much by parental education.

Finally, Figure 7a shows that the wage gain is consistent across both short- and long-term horizons post-graduation. Figure 7b indicates that the effects are most pronounced in Economic Sciences, Education, and Engineering, but not in Health or Law. Several reasons could explain the lack of response in these fields.



(a) Horizon after Graduation (b) Academic Area

Figure 7: Wage Effect Heterogeneity: Without compositional changes

Note: This figure presents Dynamic Difference-in-Differences estimates of the effect of the exam-to-project policy on wages, excluding compositional changes. Panel (a) examines the wage effect across different post-graduation time horizons. Panel (b) explores heterogeneity across academic areas. The solid horizontal line represents a zero effect for reference.

First, we confirm that these areas had similar exam exposure to other programs, ruling out the absence of treatment as a factor. However, both areas involve board certification exams taken after graduation. That said, ULACIT's Health programs do not include Medicine—the only health field requiring such certification—and the Law Bar Exam was introduced many years after the policy change and even falls outside the scope of our last admission cohort used in the main estimations. As a result, these external exams are unlikely to influence the observed patterns. The lack of wage gains in these fields is likely due to the absence of the mechanisms driving gains in other disciplines.

Students in these fields may not be able to engage in profession-related work before graduation and certification, particularly in Health, due to the sensitive nature of their work. Moreover, many doctors (and to some extent, lawyers) work within a single large entity—the Public Health System for doctors and the Judiciary System for lawyers. Since each of these sys-

tems function as unified institutions or firms, there is no meaningful change in the number of coworkers from the same university, limiting the impact of alumni networks. Similarly, the better-paying firm mechanism is less relevant, as professionals in these fields are already employed within a single, dominant employer.

Discussion on External Validity

While our empirical setting does not allow us to provide evidence on external validity, in this section we shortly explore, in an informal and speculative manner, the following questions:

Can These Findings Extend to Other Countries? While the magnitude of effects may differ, several mechanisms behind ULACIT's exam-to-project reform—such as improved soft skills and early job market alignment—are plausibly relevant in other countries. Given that U.S. students already have greater work experience during college, the gains from project-based grading may be smaller. However, the collaborative nature of projects may still enhance non-cognitive skills valued by employers. As such, while similar reforms in the U.S. would likely yield smaller wage effects than those observed at ULACIT, meaningful benefits could still arise from improved soft skills and better labor market matching.

Is the Impact of the Policy Contingent on Being an Early Adopter? Several of the policy's effects—such as grade inflation, shorter program duration, and enhanced soft skills—would likely persist under broader adoption, given their intrinsic link to project-based pedagogy. However, ULACIT's status as an early adopter did amplify the observed wage gains, particularly by allowing its graduates to stand out in the labor market. If all universities adopted similar reforms firms would no longer treat project-based graduates as a distinctive group. As a result, wage premiums would likely diminish in general equilibrium, as soft skills and work experience during college become widespread rather than differentiating. While the underlying benefits of collaborative learning would remain, their labor market returns might be substantially muted once the policy becomes the norm.

Why Haven't Other Universities Adopted This Policy?

None of the other universities in our control group—or in the country—have adopted a policy fully replacing exams with projects. Ex-ante, several possible reasons could explain this:

(a) lack of awareness of the policy, (b) complexity of execution due to coordination challenges, resource constraints, or internal administrative processes, (c) confidence in the effectiveness of the current evaluation system, (d) uncertainty about potential effects, (e) resistance from faculty, among others.

We inquired with these universities on this matter, but they were hesitant to rank or identify any single reason as their official stance. However, three key insights emerged from our discussions. First, they recognize that different evaluation systems have distinct advantages and drawbacks. Second, they have strong confidence in an exam-centric system, valuing personal achievement and its ability to more accurately measure individual competencies. Third, even after reviewing the effects—both positive and negative—of ULACIT's policy, they remain firm in their decision not to assess or implement a similar approach.

8. Conclusion

This paper examines the academic and labor market consequences of eliminating traditional class exams in favor of group-based project assessments. This reflects a broader shift in education policy away from high-stakes testing and toward alternative evaluation methods that emphasize a wider range of competencies. We exploit a natural experiment at ULACIT University—one of the top universities in Costa Rica—which in 2008 replaced exams with lower-stakes, formative project evaluations designed to create a learning environment more closely aligned with labor market demands.

Using a difference-in-differences design across admission cohorts, we first document significant compositional changes following the reform: the incoming student body had stronger math skills, a greater share of students from private high schools, and higher rates of graduate school enrollment. After accounting for these shifts, we still find that the reform led to meaningful increases in graduates' wages, associated with greater work experience during college and the development of soft skills such as networking. Wage gains were most pronounced among male students, graduates from public high schools, and students with weaker academic performance in high school. Despite its benefits, the reform appears to have contributed to unintended consequences, including grade inflation, higher faculty turnover, and slower growth in undergraduate enrollment. Importantly, we cannot directly measure learning in the

traditional, test-based sense. However, the overall pattern of results suggests that even if such learning were negatively affected, the development of other competencies more than compensated for it. As Jackson (2018) has argued, what exams miss may matter as much as what they measure. The fact that earnings effects remain strongly positive indicates that the reform left graduates better off in aggregate.

These findings highlight the transformative potential of project-based learning in higher education, suggesting it may improve labor market outcomes. At the same time, they underscore the complexities and trade-offs inherent in moving away from traditional evaluation systems—especially for institutions striving to balance educational equity, academic rigor, and labor market alignment. Taken together, our estimates capture the total impact of the reform bundle—direct changes in evaluation methods along with the compositional and institutional responses it set in motion. These results should not be interpreted as a case for eliminating exams, but rather as evidence that skills under-incentivized by exam-centered systems are strongly valued in the labor market.

9. References

Ainsworth, Aaron, Emily Penner, and Yujia Liu. 2024. "Less is More? The Causal Effect of Four-Day School Weeks on Employee Turnover." *Annenberg Institute* EdWorkingPaper No. 24-1035.

Alfaro-Ureña, Alonso, Benjamin Faber, Cecile Gaubert, Isabela Manelici, and Jose Vasquez. 2023. "Responsible Sourcing? Theory and Evidence from CostaRica." *American Economic Review* Revise and Resubmit.

Alfaro-Ureña, Alonso, Isabela Manelici, and Jose Vasquez. 2022. "The Effects of Joining Multinational Supply Chains: New Evidence from Firm-to-Firm Linkages." *The Quarterly Journal of Economics* 137 (3): 1495–1552.

Alvarez, Fernando, David Argente, Francesco Lippi, Esteban Mendez, and Diana VanPatten. 2023. "Strategic Complementarities in a Dynamic Model of Technology Adoption: P2P Digital Payments." *American Economic Review* Revise and Resubmit.

Barron, Manuel, Pia Basurto, Martin Carbajal, and Roman Zarate. 2024. "School Effects on College Outcomes in the Absence of Standardized Tests: The Role of Reputation vs. Effectiveness." *Working Paper.*

Berry, James, Harini Kannan, Shobhini Mukherji, and Marc Shotland. 2020. "Failure of frequent assessment: An evaluation of India's continuous and comprehensive evaluation program." *Journal of Development Economics* 143 (102406).

Borghesan, Emilio. 2023. "The Heterogeneous Effects of Changing SAT Requirements in Admissions: An Equilibrium Evaluation." *Journal of Political Economy* Revise and Resubmit.

Bowden, Brooks, Viviana Rodriguez, and Zach Weingarten. 2023. "The Unintended Consequences of Academic Leniency." *Working Paper*, https://www.edworkingpapers.com/sites/default/files/ai23-836.pdf.

Busso, Matías, Sebastian Montaño, and Juan Muñoz. 2023. "Signaling Specific Skills and the Labor Market of College Graduates." *IZA Discussion Paper No. 16449*, https://doi.org/http://dx.doi.org/10.2139/ssrn.4572657.

Chetty, Raj, David Deming, and John Friedman. 2023. "Diversifying Society's Leaders? The Determinants and Causal Effects of Admission to Highly Selective Private Colleges." *NBER Working Paper 31492*, https://doi.org/10.3386/w31492.

Clark, Damon, and Paco Martorell. 2014. "The Signaling Value of a High School Diploma." *Journal of Political Economy* 122 (2): 282–318.

Conare. 2017. "National State of the Education in Costa Rica: Chapter 5, The Evolution of Higher Education."

———. 2023. "National State of the Education in Costa Rica: Chapter 5, The Evolution of Higher Education."

Crepon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot, and Philippe Zamora. 2013. "Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment." *The Quarterly Journal of Economics* 128 (2): 531–580.

Dang, Haianh, Paul Glewwe, Jongwook Lee, and Khoa Vu. 2022. "The Impact Evaluation of Vietnam's Escuela Nueva (New School) Program on Students' Cognitive and Non-cognitive Skills." *IZA DP No. 15005*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4114534.

Deming, David. 2017. "The Growing Importance of Social Skills in the Labor Market." *The Quarterly Journal of Economics* 132 (4): 1593–1640.

Deming, David, and Lisa Kahn. 2018. "Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals." *Journal of Labor Economics* 36 (S1).

Deming, David, and Ben Weidmann. 2021. "Team Players: How Social Skills Improve Team Performance." *Econometrica* 89 (6): 2637–2657.

Diamond, Rebecca, and Petra Persson. 2017. "The Long-term Consequences of Teacher Discretion in Grading of High-stakes Tests." *Working Paper*, https://web.stanford.edu/~diamondr/DiamondPersson.pdf.

Ebenstein, Avraham, Victor Lavy, and Sefi Roth. 2016. "The Long-Run Economic Consequences of High-Stakes Examinations: Evidence from Transitory Variation in Pollution." *American Economic Journal: Applied Economics* 8 (4): 36–65.

Eren, Ozkan, and Daniel Henderson. 2008. "The impact of homework on student achievement." *Econometrics Journal* 11 (2): 326–348.

———. 2011. "Are we wasting our children's time by giving them more homework?" *Economics of Education Review* 30 (5): 950–961.

Felegi, Brianna. 2024. "The Impacts of Removing College Entrance Exams: Evidence from the Test-Optional Movement." *Revise and Resubmit Southern Economic Journal.*

Feng, Andy, and Georg Graetz. 2017. "A Question of Degree: The Effects of Degree Class on Labor Market Outcomes." *Economics of Education Review* 61:140–161.

Fischer, Alexander, Andrei Gorshkov, Trondur Moller, and Jeanette Walldorf. 2023. "Peers, Heirs, and Careers: Labor Market Effects of Alumni Networks." *Working Paper.*

Freier, Ronny, Mathias Schumann, and Thomas Siedler. 2015. "The earnings returns to graduating with honors — Evidence from law graduates." *Labour Economics* 34:39–50.

Friedman, John, Bruce Sacerdote, and Michele Tine. 2024. *Standardized Test Scores and Academic Performance at Ivy-Plus Colleges.* Opportunity Insights. https://opportunityinsights.org/wp-content/uploads/2024/01/SAT_ACT_on_Grades.pdf.

Grodner, Andrew, and Nicholas Rupp. 2013. "The Role of Homework in Student Learning Outcomes: Evidence from a Field Experiment." *The Journal of Economic Education* 44 (2): 93–109.

Heissel, Jennifer, Emma Adam, Jennifer Doleac, David Figlio, and Jonathan Meer. 2021. "Testing, Stress, and Performance: How Students Respond Physiologically to High-Stakes Testing." *Education Finance and Policy* 16 (2): 183–208.

ILO. 2018. *Women and Men in the Informal Economy: A Statistical Picture (Third Edition).* International Labour Office. https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/documents/publication/wcms_626831.pdf.

INEC. 2015. "National Institute of Statistics and Census: Continuous Employment Survey on the Informal Sector."

Jackson, Kirabo. 2018. "What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes." *Journal of Political Economy* 126 (5).

Johnston, Neil. 2025. "Oxford and Cambridge to Move Away from Traditional Exams to Boost Results of Minorities." Accessed January 25, 2025. https://www.telegraph.co.uk/news/2025/01/25/oxford-cambridge-move-away-traditional-exams-boost-results/.

Kalenkoskia, Charlene, and Sabrina Pabilonia. 2016. "Does high school homework increase academic achievement?" *Education Economics* 25 (1): 45–59.

Khoo, Pauline, and Ben Ost. 2018. "The effect of graduating with honors on earnings." *Labour Economics* 55:149–162.

Kingston, Sally. 2018. "Project Based Learning Student Achievement: What Does the Research Tell Us?" *PBL Evidence Matters* 1 (1): 1–11.

Kokotsaki, Dimitra, Victoria Menzies, and Andy Wiggins. 2016. "Project-based learning: A review of the literature." *Improving Schools* 19 (3): 267–277.

Krajcik, Joseph, Barbara Schneider, Emily Miller, Ichien Chen, Lydia Bradford, Quinton Baker, Kayla Bartz, et al. 2023. "Assessing the Effect of Project-Based Learning on Science Learning in Elementary Schools." *American Educational Research Journal* 60 (1): 70–102.

LeBarbanchon, Thomas, Diego Ubfal, and Federico Araya. 2023. "The Effects of Working While in School: Evidence from Employment Lotteries." *American Economic Journal: Applied Economics* 15 (1): 383–410.

Lincove, Jane, Catherine Mata, and Kalena Cortes. 2022. "A Bridge to Graduation: Post-Secondary Effects of an Alternative Pathway for Students Who Fail High School Exit Exams." *Bureau of Economic Research, No. w29742.*

Mendez, Esteban, and Diana VanPatten. 2024. "Voting on a Trade Agreement: Firm Networks and Attitudes Toward Openness." *Review of Economic Studies, accepted* NBER Working Paper No. 30058.

Pietro, Giorgio. 2017. "Degree classification and recent graduates' ability: Is there any signalling effect?" *Journal of Education and Work* 30 (5): 501–514.

Rambachan, Ashesh, and Jonathan Roth. 2023. "A More Credible Approach to Parallel Trends." *The Review of Economic Studies* 90 (5): 2555–2591.

Rampell, Catherine. 2011. "A History of College Grade Inflation." Accessed March 22, 2024. https://archive.nytimes.com/economix.blogs.nytimes.com/2011/07/14/the-history-of-college-grade-inflation/.

Roth, Jonathan. 2022. "Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends." *AER: Insights* 4 (3): 305–322.

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

Rothstein, Jesse. 2004. "College Performance Predictions and the SAT." *Journal of Econometrics* 121 ((1-2)): 297–317.

Saavedra, Anna, Kari Morgan, Ying Liu, Marshall Garland, Amie Rapaport, Alyssa Hu, Danial Hoepfner, et al. 2022. "The Impact of Project-Based Learning on AP Exam Performance." *Educational Evaluation and Policy Analysis* 44 (4): 638–666.

Schwerdt, Guido, and Ludger Woessmann. 2017. "The information value of central school exams." *Economics of Education Review* 56:65–79.

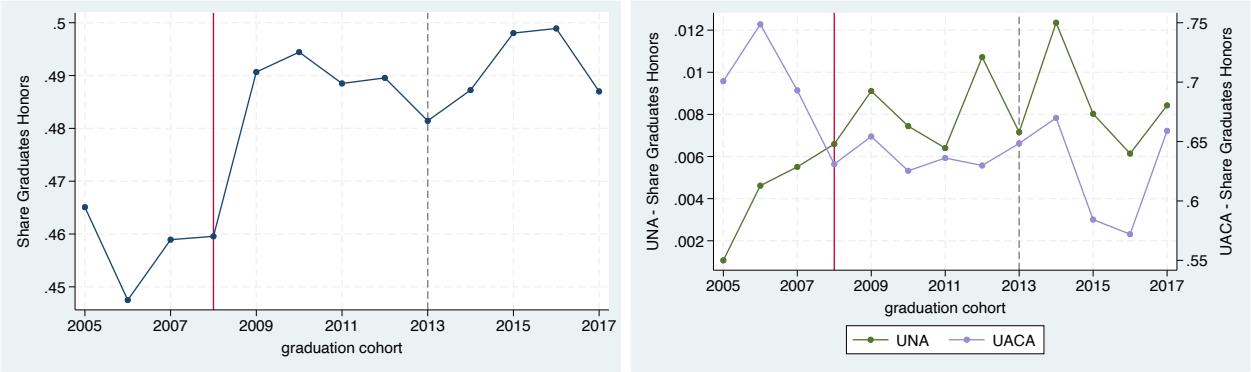
Sun, Liyang, and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics* 225 (2): 175–199.

Tucker, Jill. 2025. "SFUSD kills controversial grading proposal after backlash." Accessed May 28, 2025. https://www.sfchronicle.com/bayarea/article/sfusd-kills-controversial-grading-proposal-20349399.php.

UC. 2020. *Report of the UC Academic Council Standardized Testing Task Force (STTF).* University of California System.

Zarate, Roman. 2023. "Uncovering Peer Effects in Social and Academic Skills." *American Economic Journal: Applied Economics* 15 (3): 35–79.

10. Appendix



(a) ULACIT (b) UNA-UACA

Figure 8: Share of Graduates with Honor Distinction

Note: This figure presents the share of graduates who earned an honor distinction across different universities: (a) the share of ULACIT graduates receiving honors over time; (b) the share of graduates receiving honors at UNA and UACA, two untreated universities, over the same period. The solid vertical line in 2008 marks the introduction of the new evaluation system without exams, and the corresponding dotted line indicates the point when most graduating students at ULACIT were fully exposed to the system.

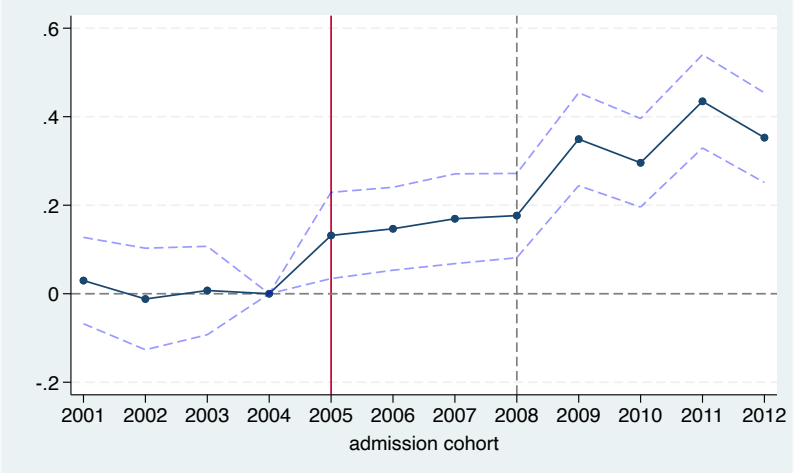


Figure 9: Wage Effect: Excluding majors which benefited from CIT

Note: This figure presents Dynamic Difference-in-Differences estimates of the effect of the exam-to-project policy on wages, excluding majors that benefited from the 2013 Technology Center infrastructure project. The solid vertical line in 2005 marks the introduction of the new evaluation system without exams, and the corresponding dotted line marks the first cohort of students who enrolled at ULACIT with full knowledge of the exam-to-project policy rollout and thus received full exposure to the system.

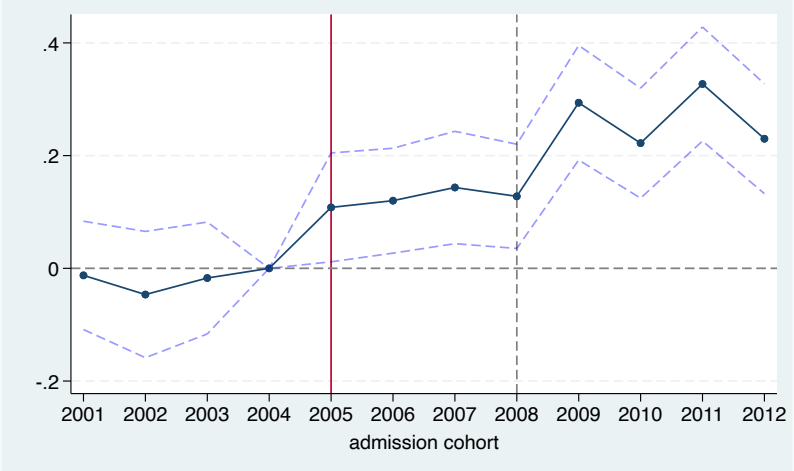
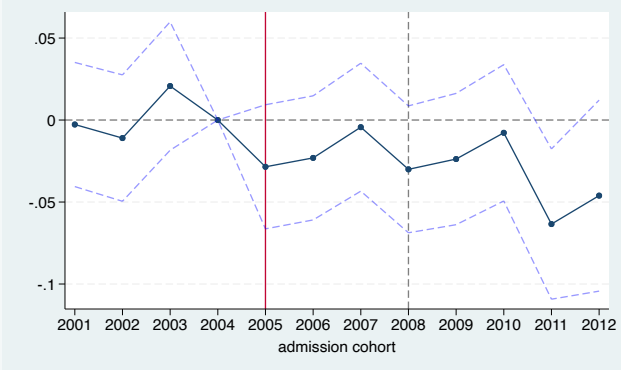
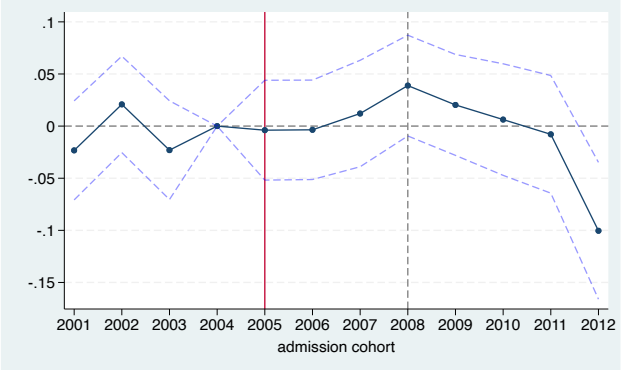


Figure 10: Wage Effect: Adding Year Fixed Effects and Labor Experience Controls

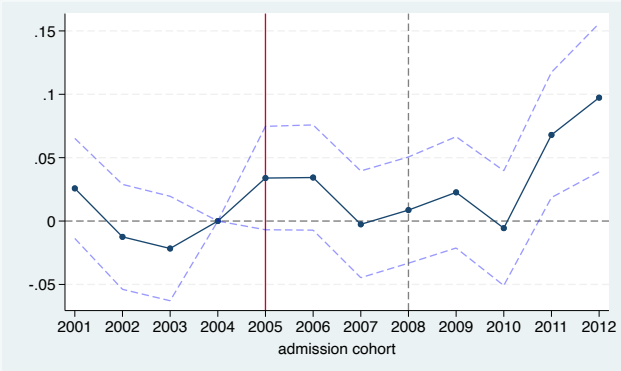
Note: This figure presents Dynamic Difference-in-Differences estimates of the effect of the exam-to-project policy on wages, controlling for calendar year fixed effects and cumulative (time-varying) labor market experience, measured in years. The solid vertical line in 2005 marks the introduction of the new evaluation system without exams, and the corresponding dotted line marks the first cohort of students who enrolled at ULACIT with full knowledge of the exam-to-project policy rollout and thus received full exposure to the system.



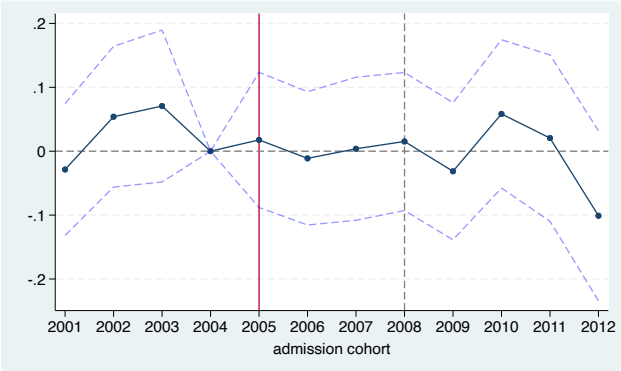
(a) UCR Pseudo-Treated



(b) TEC Pseudo-Treated



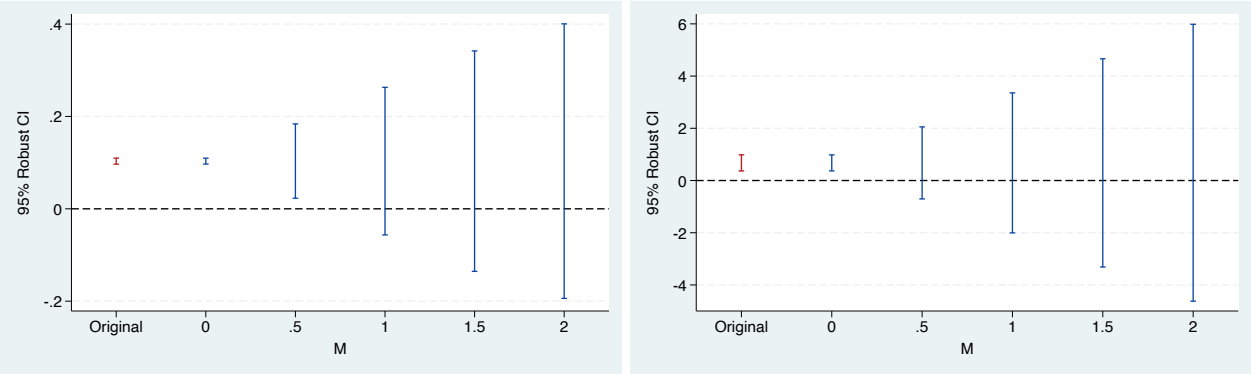
(c) UNA Pseudo-Treated



(d) UACA Pseudo-Treated

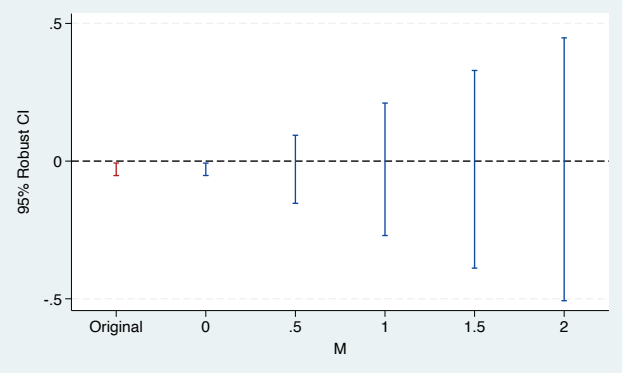
Figure 11: Wage Effect: Placebo Test on Pseudo-Treated Control Universities

Note: This figure presents Dynamic Difference-in-Differences estimates of the effect of the exam-to-project policy on labor market outcomes. Each panel conducts a placebo test by assigning one control university a pseudo-treated status while using the remaining three as controls; ULACIT is excluded from the sample in all cases. Panel (a) assigns treatment to UCR, Panel (b) to TEC, Panel (c) to UNA, and Panel (d) to UACA. The solid vertical line in 2005 marks the introduction of the new evaluation system without exams, and the corresponding dotted line marks the first cohort of students who enrolled at ULACIT with full knowledge of the exam-to-project policy rollout and thus received full exposure to the system.

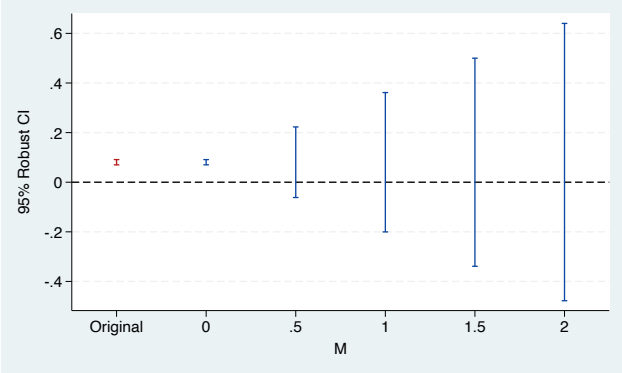


(a) Share Enrollment Grad Program

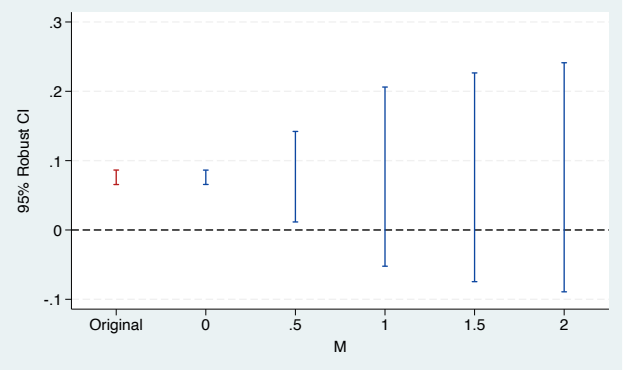
(b) HS math exit-GPA



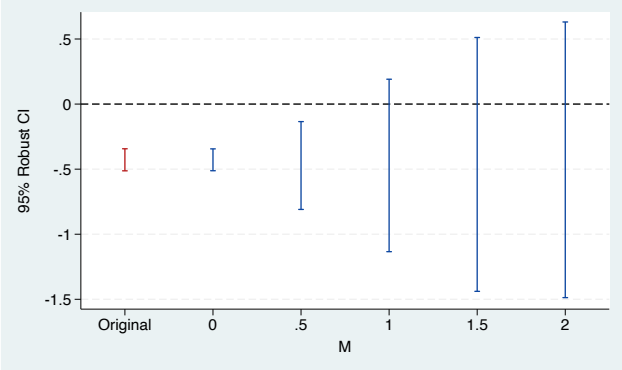
(c) Share of Male



(d) Share of Private HS



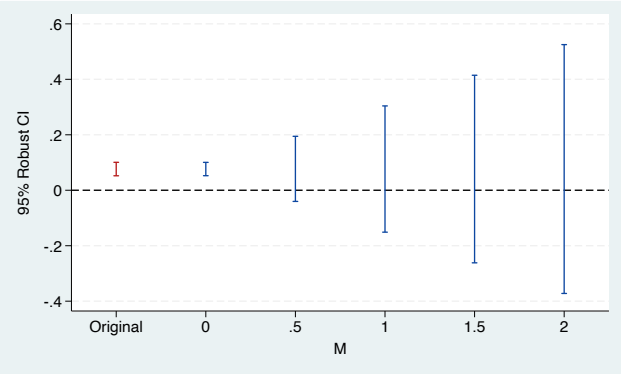
(e) Share of Faculty Exit



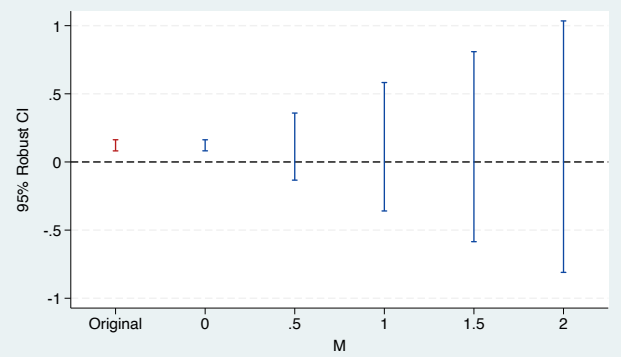
(f) Faculty Wage

Figure 12: HonestDiD Sensitivity Bounds (Relative Magnitudes M) for Bundled Post-Treatment Effect: Student Composition Outcomes and Faculty Responses

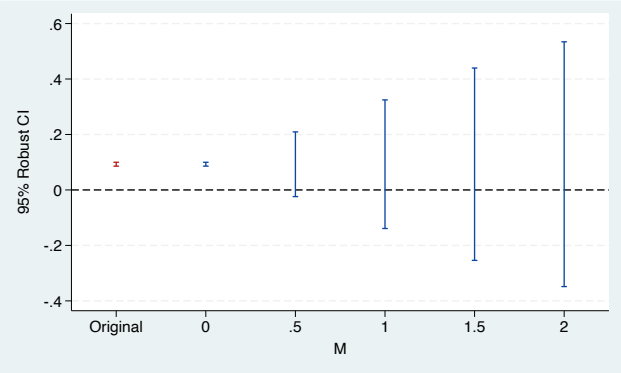
Note: The table reports HonestDiD robust confidence intervals for the bundled post-treatment effects on the main student composition outcomes and faculty responses under relative-magnitude restrictions (Rambachan and Roth 2023). The bundled effect is the equal-weighted average of post-treatment event-time coefficients. The relative-magnitude restriction imposes that the post-treatment violation of parallel trends is no larger than M times the maximum pre-treatment violation. We report 95% robust CIs for $M \in \{0, 0.5, 1, 1.5, 2\}$. The row labeled "Original" shows the conventional 95% CI under standard parallel-trends assumptions.



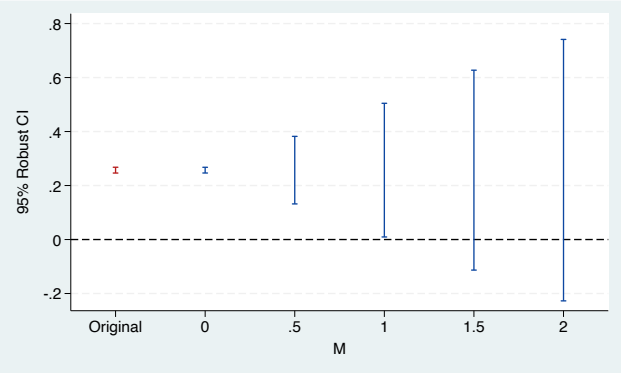
(a) Connections normalized by Major Cohort size



(b) Work before Graduation



(c) Formal Work



(d) Wage

Figure 13: HonestDiD Sensitivity Bounds (Relative Magnitudes M) for Bundled Post-Treatment Effect: Soft Skills and Labor Outcomes

Note: The table reports HonestDiD robust confidence intervals for the bundled post-treatment effects on the main soft skills and labor outcomes under relative-magnitude restrictions (Rambachan and Roth 2023). The bundled effect is the equal-weighted average of post-treatment event-time coefficients. The relative-magnitude restriction imposes that the post-treatment violation of parallel trends is no larger than M times the maximum pre-treatment violation. We report 95% robust CIs for $M \in \{0, 0.5, 1, 1.5, 2\}$. The row labeled "Original" shows the conventional 95% CI under standard parallel-trends assumptions.