

Improving Student Outcomes with Deadlines: Evidence from the German Higher Education System

By ZOUHIER KASSABALLI*

Higher education systems struggle with low academic productivity, as many students progress slowly or fail to complete their degrees. In highly flexible institutional environments, the absence of short-term constraints can unintentionally weaken incentives for sustained effort and reinforce procrastination. This study exploits a natural experiment at a German university to test whether limiting this flexibility through course-completion deadlines improves outcomes. Two study programs introduced deadlines in 2016: one early and targeted (first-semester), the other delayed and comprehensive (second-year). Using administrative panel data and a difference-in-differences approach, I find that the early deadline led to significant improvements, increasing earned credits by 18 percent in the first year and boosting long-term graduation rates by 7 percentage points. The delayed deadline generated similar effect magnitudes—12% increase in long-term earned credits and a 5.7 percentage-point rise in on-time graduation—though these estimates lacked statistical precision in the primary specification and reached significance only after propensity-score matching improved covariate balance. Synthetic difference-in-differences estimates further corroborate these findings, yielding statistically significant results of similar magnitude. Importantly, dropout rates and GPA remained unchanged, indicating that improvements are due to increased effort rather than negative selection. Lastly, consistent effects across ability levels suggest that deadlines serve as a universal pacing tool, showing that institutions can intentionally generate academic momentum through policy.

JEL Codes: I21, I23, D91, C93

Keywords: higher education, deadlines, natural experiment, academic momentum

* University of Erfurt (zouhier.kassaballi@uni-erfurt.de).

1. Introduction

A central question in the design of higher education institutions is how to structure academic environments to maximize human capital accumulation. Theoretically, institutional flexibility allows students to allocate their effort over time optimally; in practice, however, behavioral frictions often disrupt this optimization (Lavecchia & Oreopoulos, 2016). In Germany, these frictions produce sizable inefficiencies: roughly 30% of bachelor's students drop out after an average of 4.7 semesters, and only about one-third of graduates finish on time (Heublein et al., 2017; Statistisches Bundesamt, 2024). These outcomes are often linked to the system's strong emphasis on student autonomy, reflected in tuition-free enrollment and relatively flexible examination regulations (Bäulke and Dresel, 2023). Lacking immediate constraints, students face little cost for delay, making them structurally vulnerable to drifting off track.

Mandatory course-completion deadlines target these behavioral frictions by serving as institutional commitment devices. By attaching immediate consequences to inaction, they force students to internalize future costs and sustain effort (Ariely & Wertenbroch, 2002; Kaur et al., 2015). However, the empirical evidence of such constraints is mixed: while some interventions successfully increase effort (Ariely & Wertenbroch, 2002), others fail to improve completion rates or yield null results (Burger et al., 2011). This inconsistency highlights a critical trade-off: while deadlines can generate momentum, they also impose rigidity that may disadvantage students who rely on flexibility to balance external obligations. Consequently, the effectiveness of deadlines likely depends on the specific balance among timing, scope, and enforcement strictness.

In response to sub-optimal student outcomes, universities have begun experimenting with mandatory course-completion deadlines to enforce consistent engagement. I study two such reforms introduced in 2016 at a German university of applied sciences, where two large study programs adopted course-completion deadlines while other programs continued under unchanged regulations. The Business Administration (BA) program introduced strict first-semester deadlines for two foundational modules. In contrast, the Social Work (SW) program required students to complete all first-year courses by the end of the second year. This design variation provides analytic leverage to assess whether timing (early versus delayed) or scope (targeted versus comprehensive) matters more for effectiveness.

I exploit administrative panel data covering over 11,500 students across seven cohorts (2012–2018) and estimate difference-in-differences models with program and cohort fixed effects. To

address the small-cluster inference problem inherent in settings with few treated units, I implement randomization inference (RI-t) following MacKinnon and Webb (2018, 2020). As additional robustness layers, I re-estimate the effects using propensity-score matching (PSM) to improve covariate balance and synthetic difference-in-differences (SDID) (Arkhangelsky et al., 2021) to construct optimized synthetic counterfactuals, ensuring the findings are robust to demographic compositional shifts and potential trend heterogeneity.

The results show that mandatory deadlines can meaningfully shift academic trajectories, particularly when imposed early and with clear consequences. In BA, students entered their programs with greater momentum, completing substantially more coursework by the time the deadline first bound. In the first year, treated cohorts earned roughly 7 additional credit points (approximately 1.4 additional courses), an 18% increase over pre-reform cohorts. These short-run gains are statistically significant across multiple inference methods (RI-t $p = 0.001$) and were achieved without any decline in GPA and without an increase in dropout rate. These results indicate that students responded to the reform by raising total study effort, not by diluting performance or exiting early.

This early acceleration translated into sustained long-run improvements. By semester seven, treated students had accumulated 21.9 additional credits, a 17% gain that is marginally significant under randomization inference (RI-t $p=0.054$) and strongly significant in propensity-score-matched and SDID specifications (bootstrap $p = 0.001$). This represents a proportional increase nearly identical to the short-run effect, demonstrating that the reform initiated a self-reinforcing trajectory rather than merely shifting effort forward in time. Cumulative dropout fell by 5.5 percentage points (from 26% to 20%). While this estimate is not statistically significant under conservative inference methods (RI-t $p = 0.336$), it becomes marginally significant in propensity-score matched specifications ($p=0.058$). Importantly, it suggests that improved outcomes were achieved in larger and more heterogeneous cohorts, ruling out mechanical selection effects. Moreover, GPA remained stable, confirming students maintained quality while accelerating pace.

Finally, the reforms substantially enhanced degree completion rates. In the primary specification, graduation within nine semesters rose by approximately 7 percentage points ($p = 0.07$ under randomization inference). These gains prove robust across alternative identification strategies; point estimates remain stable and reach standard significance levels when estimated via PSM ($p = 0.027$) and SDID ($p = 0.03$) approaches. Similarly, on-time graduation increased by

roughly 4.5 percentage points, with precise estimates emerging in both the PSM and SDiD approaches. By resolving the statistical imprecision inherent in the baseline model, these estimators provide consistent evidence that mandatory deadlines improved graduation efficiency. The persistence of these results suggests that deadlines do more than prompt short-run compliance; they durably reshape students' academic trajectories.

The SW deadline regime generated improvements similar in magnitude to the BA effects but with notably greater statistical uncertainty in the primary specification. In the unadjusted difference-in-differences specification, treated students earned roughly 7.8 additional credits by semester four and 12.4 additional credits by semester seven, representing gains of about 10% and 9% relative to pre-reform cohorts. Although the estimated effects move consistently in the expected direction, they do not reach statistical significance under conservative randomization-inference procedures (RI-t $p > 0.20$). This lack of significance reflects the limited precision inherent in a setting with only one treated program and demographic differences between the treated and control groups.

To improve comparability and mitigate statistical noise, I re-estimate the effects using PSM and SDiD specifications. Both methods yield convergent evidence of substantial effects: the matched specification estimates 10.1 additional credits by semester four ($p = 0.017$) and 17 credits by semester seven ($p = 0.015$). The synthetic DiD framework corroborates this momentum, yielding significant gains in long-term credits (11.5 credits, $p = 0.002$) and a 7.1 percentage-point increase in nine-semester completion ($p = 0.010$). The close alignment of point estimates across specifications suggests that baseline imprecision reflected power constraints rather than a null response, though conclusions for Social Work require greater caution than for BA, given the heavier reliance on modeling assumptions.

Evidence on heterogeneous treatment effects shows that the benefits of deadlines are broadly shared rather than concentrated among any particular ability group. In BA, students in the top and middle terciles of the pre-university GPA distribution register clear and statistically significant gains in credit accumulation and timely degree completion (bootstrap $p < 0.10$). Students in the lowest tercile exhibit effects of similar or even larger magnitude, but the estimates are noticeably less precise (bootstrap $p > 0.23$), reflecting limited statistical power rather than a qualitatively different response. Triple-difference tests formally corroborate this pattern, yielding no evidence of differential effects across ability groups (bootstrap $p > 0.22$). The uniform pattern of effects

indicates that deadlines do not selectively assist weaker students but are broad tools that benefit students across the ability spectrum.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 describes the German higher education system and the institutional setting of the university under study. Section 4 details the 2016 policy reforms. Section 5 presents administrative data and descriptive evidence. Section 6 explains the empirical strategy, including the inference procedures. Section 7 reports the main results, covering short- and long-run effects, event-study evidence, and heterogeneous treatment effects by student ability. Section 8 provides robustness checks using propensity-score–matched difference-in-differences estimators. While Section 9 implements synthetic difference-in-differences to optimize pre-treatment fit on outcomes themselves. Section 10 discusses the findings and concludes.

2. Literature Review

Deadlines as Incentives—Students often delay costly effort because the benefits of studying lie in the future while the costs are immediate, creating predictable time-inconsistency and procrastination (O'Donoghue & Rabin, 1999; Akerlof, 1991). In environments with weak external constraints, this mismatch leads to underinvestment in effort. Deadlines can mitigate these failures by raising the short-run cost of delay and helping individuals commit to effort schedules they struggle to sustain voluntarily. (Ariely & Wertenbroch, 2002).

Empirical work confirms that deadlines can increase timely effort. Ariely and Wertenbroch (2002) show that externally imposed, evenly spaced deadlines improve performance relative to flexible schedules or self-imposed deadlines. Deadlines act through heightened loss aversion and salient reference points (Zamir, Lewinsohn-Zamir & Ritov, 2017) and reduce psychological distance to future costs, generating goal-gradient responses. Evidence outside education shows similar dynamics: workers voluntarily accept penalty-based deadlines that concentrate effort (Kaur, Kremer & Mullainathan, 2015).

In the German higher-education system, Ostermaier (2018) shows that removing a binding deadline for introductory modules led students to strategically postpone effort. When the deadline was abolished, students increasingly submitted blank exams, thereby extending their study duration. The response was most pronounced among mid-ability students, illustrating that even modest institutional structure can discipline time-inconsistent study behavior in highly flexible

environments. However, design matters: excessively rigid or poorly timed deadlines can reduce completion (Bisin & Hyndman, 2020). This paper examines institutional course-completion deadlines, a relatively unexplored policy tool distinct from assignment deadlines, and tests whether early versus delayed designs generate different behavioral responses.

Academic Dismissal Policies: Incentives versus Selection—A second related strand of research examines academic dismissal policies (AD), which set minimum first-year performance thresholds that students must meet to remain enrolled. A central question in this literature is whether such rules primarily change behavior (incentives) or merely alter which students persist (selection). The evidence is mixed, even within the Dutch higher-education context where most studies are conducted.

Several studies conclude that AD policies function primarily as selection mechanisms. Arnold (2015) reports that while credit thresholds increased early attrition by 6–7 percentage points and accelerated the progress of survivors, they left the four-year completion rate for the starting cohort unchanged. This indicates that in some settings, dismissal simply brings dropout forward in time without generating net gains in human capital. Cornelisz et al. (2020) provide the most direct evidence of this selection channel: using a regression discontinuity design, they show that students just above and below the dismissal cutoff exhibit identical long-run graduation rates, implying that the dismissal threat itself did not causally alter academic trajectories. Sneyers and De Witte (2017) offer a slightly more positive view on efficiency, finding that while AD policies increased first-year dropout by roughly 7.5 percentage points, they also led to higher graduation rates among reenrolled students, suggesting that rigorous selection can improve success rates for those who persist.

Other studies examining comparable Dutch policies reach opposite conclusions. Schmidt et al. (2022) find that first-year credit requirements increased first-year without raising dropout, clear evidence of substantial effort responses. Effects concentrated among the weakest students, with completion rates for the lowest-GPA group rising from 17% to 48%. Arnold (2023) similarly shows that tightening performance standards reduces procrastination and increases credit accumulation among low-prepared students. These findings suggest AD rules can induce genuine behavioral responses within the same institutional environment where other studies find only selection. This paper extends this literature by providing causal evidence from the German higher

education system, analyzing whether targeted, early milestones can trigger similar incentive effects without the broad-scope intensity of full-curriculum dismissal policies.

Academic Momentum and Long-Run Impacts—The long-term effects examined in this study relate to the theory of academic momentum, which posits that early academic success generates self-reinforcing trajectories toward degree completion (Attewell et al., 2012). Empirical research shows that intensive early engagement—such as enrolling in a full course load during the first semester—raises graduation probabilities by 6 to 10 percentage points (Belfield et al., 2016; Attewell & Monaghan, 2016), whereas early disengagement, such as part-time enrollment in the first semester, reduces completion rates by 5 to 13 percentage points (Attewell et al., 2012). These findings suggest that early achievement builds academic human capital, strengthens study habits, and reinforces institutional commitment.

However, most existing evidence relies on observational data or student-driven choices, leaving it unclear whether institutions can actively induce momentum through policy. There is limited evidence on whether requiring students to adopt a faster pace generates the same benefits as voluntarily choosing one. This study addresses that gap by testing whether imposing external deadlines can trigger this positive feedback loop. Thereby, examining if short-term compliance translates into durable gains in credit accumulation, persistence, and on-time graduation.

3. Institutional Setting

3.1 The German Higher Education System

Germany's higher education system has two main types of institutions: research-oriented universities and practice-focused universities of applied sciences (UAS). Together, they enroll around 2.9 million students, with UAS institutions accounting for approximately 38% of all bachelor's enrollments (Statistisches Bundesamt, 2024). German universities organize their programs under the European Credit Transfer and Accumulation System (ECTS), which standardizes academic progression across Europe. This framework outlines a clear path to timely graduation, recommending that full-time bachelor's students complete 30 credits per semester.

In reality, few students adhere to this pace. Among graduates nationwide in 2023, the average time to degree was 8.5 semesters. Only 29.8% graduated on time, while 41.8% needed up to an extra year, and 28.4% took three or more additional semesters (Statistisches Bundesamt, 2024).

Beyond these delays, dropout rates are high: approximately 30% of students abandon their studies, typically after 4.7 semesters (Heublein et al., 2017). These patterns reveal substantial inefficiencies in academic progression and student-program matching.

Three institutional characteristics contribute to these inefficiencies. First, tuition fees are mostly subsidized, which reduces the financial burden of prolonged enrollment¹. Second, academic culture places strong emphasis on student autonomy and self-regulation, with few binding deadlines or continuous assessments (Bäulke & Dresel, 2023). Third, examination policies allow students to defer or repeat exams multiple times, sometimes without limit, enabling postponement of academic obligations. Together, these features create an environment where delaying progress carries little immediate cost, making procrastination particularly likely.

3.2 The Institutional Context

The policy reform analyzed in this paper was implemented at a large UAS in Germany. UAS institutions form a core component of the German tertiary system, providing practice-oriented education that integrates theoretical training with professional application. The studied university, enrolling roughly 12,000 students across more than twenty bachelor's programs, is representative of this sector. Its structure is centrally administered but with program-level autonomy over examination and progression policies, which creates within-institution variation ideal for identifying how academic policies influence student outcomes.

Study Programs—The analysis focuses on two large programs: BA and SW, which together account for 18% of university enrollment. Both are nationally significant. Business Administration is one of Germany's most popular fields, accounting for 8.5% of first-year students, while Social Work accounts for 4.1% of total enrollment (121,664 students nationwide in 2024/2025). Both programs belong to the law, economics, and social sciences category, the largest field group in German higher education, enrolling 37.9% of all students (Statistisches Bundesamt, 2024). Both programs follow a structured seven-semester curriculum consistent with the ECTS framework, including a mandatory practical semester to link theory with workplace experience. While these programs had pre-existing academic progression rules (see section 4), the 2016 reform introduced significantly stricter deadlines, sharply reducing student flexibility.

¹ Public universities charge only modest administrative fees (typically around €250 per semester), and enrollment provides students with subsidized housing, healthcare, and transportation, potentially incentivizing prolonged study.

4. The Policy Reform

In 2016, two of the university's largest bachelor's programs, BA and SW, introduced course-completion deadlines through changes to their study and examination regulations. The changes were implemented quietly through unannounced regulatory amendments², minimizing any potential selection into or out of the affected programs. It is unlikely that prospective students adjusted their choices in response to the new deadlines, as detecting these changes would have required a detailed examination of program regulatory documents, a rare practice among prospective students. The following subsections provide a detailed description of each reform, with Table 1 summarizing the policy parameters.

4.1 Business Administration: Early, Targeted Deadline

Pre-2016 Policy—Before April 2016, BA students faced a single deadline: Introduction to Business had to be completed by the end of the second semester. Students were allowed two examination attempts and were automatically re-registered for the next available session following failure. Exhausting both attempts resulted in academic dismissal.

Post-2016 Policy—The April 2016 reform significantly intensified this regime, shifting toward a high-stakes, early-intervention model. This intensification occurred along two dimensions. First, the deadline was advanced from the end of the second semester to the end of the first (temporal intensification). Second, the policy's scope was expanded to include "Business Mathematics" as a second mandatory course with a deadline (scope expansion). Consequently, students were required to complete both "Introduction to Business" and "Business Mathematics" by the end of their first semester. Failure to do so triggered automatic re-registration for the second semester. Failure to pass by the second attempt (end of year one) resulted in academic dismissal³. The BA reform therefore represents an early, highly stringent, and targeted enforcement regime.

² Aside from the introduction of binding course-completion deadlines, the 2016 regulatory updates did not alter the first-year curriculum, course content, examination rules, or degree requirements. All other changes consisted of minor editorial or administrative adjustments—such as updated module titles and coding, that left academic content and workload unchanged (see Appendix Table A7 for a summary).

³ These two modules are designated as 'orientation modules' (*Orientierungsprüfung*), which carry stricter conditions than the rest of the curriculum. While these specific courses are limited to one retake (two attempts total) and subject to automatic re-registration, standard modules in the program allow for three examination attempts, do not trigger automatic re-registration, and can be deferred by the student.

4.2 Social Work: Comprehensive, Broad Deadline

Pre-2016 Policy—Before July 2016, SW students studied under a flexible system with no course-completion deadlines. Examinations could be scheduled or deferred freely across semesters. Students were allowed three exam attempts per module, with no risk of dismissal except after final failure.

Post-2016 Policy—The July 2016 reform introduced deadlines, requiring the completion of all first-year modules (60 credits) by the end of Semester 4. Examinations not completed by this deadline were automatically recorded as failed, counting as one of the three permitted attempts. Missed or failed exams were automatically re-registered for the next available session, and students who exhausted all three attempts in any module were dismissed from the program.

This moderate-intensity, delayed-intervention policy was designed to promote steady academic progress by introducing time-bound accountability while giving students sufficient flexibility to adjust during their first two years. In practice, the combination of the three-attempt rule and the annual scheduling of many modules meant that students could remain enrolled well beyond the nominal deadline. Because failed exams triggered automatic re-registration for the next available session, which sometimes did not occur until the following academic year, the dismissal process could stretch over several additional semesters.

4.3 Comparing the Two Policies

The BA and SW reforms differ along three central dimensions. Timing: BA deadlines bind in the first semester, whereas SW deadlines apply at the end of the fourth. Scope: BA targets two foundational orientation modules, while SW covers all first-year coursework. Intensity: BA operates under a two-attempt rule with immediate consequences, whereas SW provides three attempts and a longer compliance window. These design differences produce contrasting treatments—an early, targeted, high-stakes intervention in BA versus a broader, delayed, moderate-stakes intervention in SW. Evaluating each program separately enables a direct assessment of how the timing and intensity of deadlines shape student responses.

Table 1: Overview of Deadline Policies in Business Administration and Social Work

Policy Dimension	Business Administration (BA)		Social Work (SW)	
	Pre-2016	Post-2016	Pre-2016	Post-2016
Deadline Timing	Semester 2	Semester 1	None	Semester 4
Modules Covered	Intro to Business	Intro to Business; Business Math	No mandatory requirements	All first-year modules
Examination Attempts	2 attempts	2 attempts	3 attempts	3 attempts
Enforcement Mechanism	Automatic re-registration	Automatic re-registration	Flexible retake policy	Automatic re-registration

Notes: “Automatic re-registration” means that students who fail or miss an exam are automatically enrolled in the next available examination attempt. Academic dismissal occurs upon exhausting the maximum number of allowed attempts in any mandatory module. “None” indicates the absence of binding course-completion deadlines before 2016.

5. Data and Descriptives

5.1 Data Sources and Structure

The analysis uses administrative student-level data from the university's examination office, covering all bachelor's programs from 2012 to 2022. The data are structured as a student-semester panel, tracking individual academic trajectories from enrollment through graduation, exit, or the end of the observation window. For the main analysis, the sample includes cohorts entering between 2012 and 2018⁴. This window allows tracking each student for at least four years, up to 2022, ensuring sufficient observation for long-term outcomes such as graduation. Cohorts entering between 2012 and 2015 constitute the pre-reform (control) period, while those starting from 2016 onward define the post-reform (treatment) period.

The difference-in-differences design compares outcomes in two treated programs (BA & SW) to outcomes in 11 untreated study programs at the same university. The comparison group consists predominantly of STEM programs that experienced no deadline policy changes between 2012 and 2018. While these programs span different faculties, they share the same institutional environment and face common external factors affecting students, including macroeconomic conditions, labor market trends, and university-wide policies. This common framework supports the validity of the comparison group, provided pre-reform trends are parallel, a condition I test via event studies.

⁴ The 2012–2018 range was selected to maximize comparability while avoiding confounding institutional or curricular changes that occurred before 2012 or after 2018.

5.2 Outcome Variables

I examine treatment effects on both short-term and long-term academic outcomes to assess immediate behavioral responses at the point when deadlines bind and sustained momentum effects. The analysis also utilizes student background characteristics from university administrative records, including age, gender, type of high school degree, years since high school graduation, high school GPA, and citizenship status. These pre-determined covariates serve as control variables in the difference-in-differences specifications to improve precision and assess compositional trends.

Short-term outcomes are measured at the point when deadlines first bind: semester 2 for Business Administration and semester 4 for Social Work. These outcomes capture initial compliance and immediate adjustments to the new policy environment. I examine: (i) cumulative ECTS credits earned, the primary measure of academic progress; (ii) grade point average (GPA), capturing academic performance quality; (iii) cumulative dropout rate, defined as the share of enrolled students who exited the program without completing a degree; and (iv) early dropout share, the proportion of eventual dropouts who exit during the first academic year, conditional on eventually dropping out.

Long-term outcomes are measured at semester 7, the nominal program duration, to assess whether early gains translate into improved degree completion. I analyze: (i) cumulative ECTS credits, measuring sustained academic progress; (ii) GPA, testing whether quality deteriorates as students accelerate; (iii) total dropout rate, capturing persistence through the standard program duration; (iv) on-time graduation, defined as the share of students who complete their degree within seven semesters; and (v) graduation rate within nine semesters, allowing one year beyond the nominal duration⁵. Graduation outcomes are analyzed for cohorts with sufficient follow-up time, excluding the 2018 cohort from the nine-semester completion analysis.

5.3 Descriptive statistics

A key identification requirement for the difference-in-differences design is that the 2016 reforms did not trigger compositional changes in the treated programs. Table 2 compares pre-reform

⁵ Restricted to cohorts 2012–2017 to ensure a consistent nine-semester observation window. Cohorts entering after 2017 cannot be tracked for a full nine semesters due to the 2022 data limit, which would bias comparisons.

(2012–2015) and post-reform (2016–2018) cohorts in BA, SW, and the comparison programs, reporting average characteristics and DiD estimates.

Table 2: Descriptive Statistics and Trend in Students' Characteristics

	Pre-2016 (1)	Post-2016 (2)	Control 2016 (3)	Control 2016 (4)	DiD (5)
<i>Panel A: BA</i>					
High School GPA	2.532	2.446	2.776	2.705	-0.01 (0.02)
Abitur (%)	32.66	41.51	37.37	43.23	3.0 (2.22)
Female (%)	58.77	55.01	24.95	27.85	-6.65*** (2.226)
Age at Entry	21.67	21.59	21.65	21.65	-0.073 (0.142)
German (%)	96.92	96.42	97.16	95.26	1.397* (0.842)
First Uni Semester (%)	0.839	0.764	0.824	0.734	0.014 (0.018)
N	1,528	978	5,378	4,094	11,978
<i>Panel B: SW</i>					
High School GPA	2.388	2.279	2.776	2.705	-0.03 (0.02)
Abitur (%)	39.32	45.44	37.37	43.23	0.27 (2.45)
Female (%)	81.58	83.16	24.95	27.85	-1.310 (1.942)
Age at Entry	23.27	23.46	21.65	21.65	0.189 (0.265)
German (%)	97.29	96.96	97.16	95.26	1.571* (0.859)
First Uni Semester (%)	0.830	0.780	0.824	0.734	0.039** (0.020)
N	1,292	790	5,378	4,094	11,554

Notes: This table reports mean pre- and post-reform characteristics of students in treated and control programs for Business Administration (Panel A) and Social Work (Panel B). Columns (1)–(4) report group means before and after the 2016 reform. Column 5 presents the difference-in-differences estimate for each characteristic. Standard errors are reported in parentheses. High School GPA is the official secondary-school leaving grade (lower values indicate better performance). Abitur indicates possession of the academic-track university entrance certificate. Female and German refer to gender and citizenship status. Age at Entry is measured at first matriculation. First Uni Semester identifies first-time entrants.

Measures of academic preparedness show no differential changes. DiD estimates for high school GPA and the share of students holding the Abitur are statistically insignificant and close to zero, suggesting that the reform did not induce selection on ability. Demographic characteristics also align with overall trends and experience minimal differential movements.

One exception is the decline in female enrollment in BA (DiD=-6.65 pp, $p<0.01$), indicating a differential shift. This reflects divergent trends across fields: female representation decreased modestly in BA (58.8% to 55.0%) while increasing in comparison programs, which are predominantly STEM fields (24.9% to 27.9%). These opposing movements align with national patterns of declining female participation in business education alongside rising representation in STEM disciplines. Additionally, the SW program shows a small but statistically significant increase in the share of first-time university entrants (DiD = 3.9 pp). Overall, the stability of the key ability measures, high-school GPA, and Abitur share, supports the validity of the identification strategy, while the modest shifts in other characteristics mirror broader university-wide and national trends rather than responses to the reform.

6. Empirical Strategy

6.1 Difference-in-Differences Design

I estimate the causal effect of course-completion deadlines using a two-way fixed effects DiD design that compares outcome changes in treated programs (BA and SW) before and after 2016 with contemporaneous changes in 14 untreated programs at the same university. Program fixed effects absorb time-invariant differences across disciplines, such as grading cultures or inherent curriculum difficulty, while cohort fixed effects capture shocks common to each entering cohort, such as changes in macroeconomic conditions or university-wide administrative policies.

A simple pre-post comparison of treated students risks confounding deadline effects with time trends, while a cross-sectional comparison in 2016 conflates treatment effects with inherent program differences. The Difference-in-Differences (DiD) design eliminates both biases by comparing temporal changes in treated programs against those in control programs. The following specification is estimated separately for BA and SW:

$$Y_{ipt} = \alpha + \beta (Treated_p \times Post_t) + \gamma' X_i + \mu_p + \lambda_t + \varepsilon_{ipt}, \quad (1)$$

Where Y_{ipt} is the outcome of interest for student i in program p from entry cohort t . $Treated_p$ is an indicator equal to one for the treated program (e.g., BA) and zero for comparison programs. $Post_t$ is an indicator equal to one for cohorts enrolling in or after 2016. The vector X_i includes pre-determined student characteristics⁶: high school GPA, Abitur holder, gender, age at immatriculation, foreigner status, and an indicator for being in the first university semester. μ_p and λ_t represent program and cohort fixed effects, respectively. Standard errors are clustered at the program level to account for within-program correlation.

The coefficient β identifies the average treatment effect on the treated under the parallel-trends assumption. This assumption requires that, absent the reform, treated and control programs would have followed similar outcome trajectories. I formally test this assumption using event studies (see section 6.2). Finally, I complement the primary analysis with matched DiD (Section 8) and Synthetic Difference-in-Differences (Section 9). These methods strengthen causal identification by restricting comparisons to observably similar students and constructing weighted counterfactuals that mirror the treated program's pre-reform trajectory with greater precision than standard linear DiD.

Inference with Few Treated Clusters—A central challenge for inference is the small number of clusters: the analysis includes only 15 study programs, and each reform applies to a single treated program. Because the estimation clusters standard errors at the study-program level, conventional cluster-robust variance estimators are prone to over-rejection in settings with few clusters (Cameron and Miller, 2015). To ensure valid inference, I rely primarily on randomization inference (RI-t) following MacKinnon and Webb (2020). By randomly permuting treatment assignment across programs (999 replications), this method generates exact finite-sample p-values independent of asymptotic assumptions. For transparency and completeness, I also report wild cluster bootstrap p-values (Cameron, Gelbach, and Miller, 2008) and conventional program-clustered standard errors.

⁶ Zeldow and Hatfield (2021) illustrate that time-invariant covariates will not introduce bias in DiD analysis if their effect on the outcome is constant over time. Following this, time-invariant covariates are included in the model to improve precision. These covariates show similar trends across groups, and their effect on student outcomes remains stable across semesters.

6.2 Event-Study Specification

To assess the validity of the parallel-trends assumption and to trace the dynamic effects of the deadline policy across cohorts, I estimate an event-study version of the DiD model. The event-study replaces the single DiD interaction with a series of relative-cohort indicators that trace the evolution of outcomes before and after treatment adoption.

The event-study model is specified as follows:

$$Y_{\{i p t\}} = \alpha + \sum_{k=-4}^{-2} \delta_k \times Treated_{ptk} + \sum_{k=0}^2 \delta_k \times Treated_{ptk} + \gamma' X_i + \mu_p + \lambda_t + \varepsilon_{ipt}, \quad (2)$$

where k indexes cohort years relative to the policy implementation year ($2016 = 0$). $Treated_{ptk}$ is an indicator variable equal to one if treated program p in cohort t is observed k years relative to policy implementation, and zero otherwise. The first summation captures the pre-treatment coefficients for cohorts 2012 ($k = -4$) through 2014 ($k = -2$). The 2015 cohort ($k = -1$) serves as the omitted reference category. The second summation captures post-treatment effects for cohorts 2016 ($k = 0$) through 2018 ($k = 2$).

The pre-treatment coefficients δ_k for $k < 0$ test whether the treated and control programs exhibited differential trends before the policy. Under the parallel trends assumption, these coefficients should be jointly insignificant and close to zero. I formally test this using an F-test of joint significance and present the event-study estimates both graphically and in tabular form.

6.3 Heterogeneous Effects by Student Ability

Deadline policies may affect students differently depending on their prior academic preparation. To examine whether the reform generates heterogeneous responses across ability levels, I classify students into terciles based on their high-school GPA (Low, Medium, High ability). I then estimate the following fully interacted triple-difference (DDD) specification:

$$Y_{ipt} = \alpha + \beta_M(Treated_p \times Post_t) + \beta_L(Treated_p \times Post_t \times Low_i) + \beta_H(BW_p \times Post_t \times High_i) + \alpha_L Low_i + \alpha_H High_i + \gamma' X_i + \mu_p + \lambda_t + \varepsilon_{ipt}, \quad (3)$$

where β_L and β_H capture the differential treatment effects for low- and high-ability students relative to the medium-ability baseline β_M . Low_i and $High_i$ indicate the lowest and highest ability

tertiles, and the model includes the same program and cohort fixed effects and student-level controls as in the baseline DiD specification. This fully interacted triple-difference structure allows the treatment effect of the deadline policy to vary flexibly across ability groups. In the results section, I report the subgroup-specific treatment effects, obtained by re-estimating the baseline specification within ability groups, and directly present the estimated policy impact for low-, medium-, and high-ability students.

The DDD estimator requires parallel trends in ability gaps, i.e., that differences between ability groups would have evolved similarly in treated and untreated programs in the absence of the reform. I assess this assumption using an ability-specific event-study design and a joint pre-trend test of all pre-treatment interactions. These tests serve as the basis for the formal validation of the heterogeneous treatment effects (see Appendix Figures A1 and A2).

7. Results

This section presents the main empirical results in three parts. Sections 7.1–7.3 analyze the BA reform, reporting short-term effects (Section 7.1), long-term effects (Section 7.2), and event-study evidence supporting parallel pre-trends (Section 7.3). Sections 7.4–7.6 present parallel analyses for SW. Section 7.7 examines whether effects vary by student ability, testing whether deadlines disproportionately harm weaker students.

7.1 Business Administration: Short-Term Effects

The introduction of first-semester deadlines led to a substantial, immediate increase in academic effort. Table 3 reports the short-run DiD effects measured at the end of Semester 2. By the end of the first year, treated students accumulated 7.15 additional ECTS credits ($SE = 0.83$), an 18% increase over the pre-treatment mean of 40.3 credits. This effect is highly statistically significant (RI-t $p = 0.001$) and equivalent to approximately 1.4 additional courses. Event-study estimates (Figure 1) confirm that these gains emerged sharply following the 2016 reform, with no evidence of differential pre-trends.

Importantly, this credit gain occurred without compromising academic quality or triggering mass dropout. The estimated effect on GPA is negligible and statistically insignificant (coefficient = -0.01, RI-t $p = 0.93$), indicating that students expanded total study effort rather than reallocating a fixed amount of effort across a larger course load. Similarly, the reform did not increase overall

attrition. Cumulative dropout by Semester 2 remains unchanged (-0.01 , RI-t $p = 0.92$), suggesting that the additional academic pressure induced by the earlier deadline did not trigger excess exits from the program.

Finally, I examine whether the policy affected the timing of dropout. Early dropout among eventual leavers increased by 10 percentage points, though this estimate is imprecise in the baseline specification (RI-t $p = 0.21$). Propensity-score matching substantially improves precision and confirms a significant effect (Section 8.1), suggesting the reform accelerated exit timing among students who would have dropped out rather than affecting overall dropout levels.

Table 3: Short-Term Effects of Academic Dismissal Policy (Semester 2): BA

	Credits (1)	GPA (2)	Dropout (3)	Early Dropout (4)
DiD Estimate	7.15*** (0.83)	-0.01 (0.02)	-0.01 (0.01)	0.10 (0.01)
RI-t p-value	[0.001]	[0.93]	[0.92]	[0.21]
Bootstrap p-value	[0.06]	[0.58]	[0.56]	[0.18]
Pre-treatment Mean	40.27	2.57	0.13	0.522
Effect Size (%)	17.8%	-0.7%	-8.6%	19.1%
Controls	Yes	Yes	Yes	Yes
Cohort & Program FE	Yes	Yes	Yes	Yes
Observations	11,877	11,877	11,877	5,779

Notes: The table reports DiD estimates for BA, obtained by OLS, for outcomes measured at the end of Semester 2. All models include individual controls (high school GPA, Abitur type, gender, age, first-university-semester indicator, foreign citizenship), cohort fixed effects, and study-program fixed effects. Standard errors clustered at the program level (15 clusters) are reported in parentheses. Wild-cluster bootstrap p-values and RI-t p-values from randomization inference appear in brackets. Effect sizes are calculated relative to the pre-treatment mean. “Early Dropout” denotes the share of dropouts during the first academic year. Significance stars are based on RI-t p-values: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

7.2 Business Administration: Long-Term Effects

Early gains in credit accumulation translated into sustained momentum rather than a one-time intertemporal substitution of effort. Table 4 reports the effects on long-term progression and degree completion. Treated students accumulated 21.9 additional credits by Semester 7 (SE = 1.67, RI-t $p = 0.054$), a 17% gain relative to the pre-treatment mean. Notably, this proportional increase mirrors the 18% gain observed at Semester 2, demonstrating that the early acceleration initiated a self-reinforcing trajectory rather than merely shifting effort forward in time. Importantly, this credit gain is accompanied by continued stability in overall grades, as cumulative GPA remained unchanged (RI-t $p = 0.81$).

These sustained credit gains translated into higher degree completion rates, though here the evidence must be interpreted with greater caution. Graduation within nine semesters increased by 7.0 percentage points ($SE = 0.011$, $RI-t p = 0.07$), a 12% rise relative to the pre-reform mean. While the on-time graduation increases by 4.0 percentage points ($SE = 0.01$), the effect is less precisely determined ($RI-t p = 0.34$). However, the results prove robust across alternative identification strategies; both the matched DiD and synthetic DiD frameworks yield highly significant point estimates for both on-time ($p < 0.001$) and nine-semester graduation ($p = 0.033$). (see sections 8 and 9).

Lastly, the results contradict the hypothesis that strict deadlines improve averages solely by selecting out weaker students. Cumulative dropout declined by 5.5 percentage points. While this estimate is imprecise under conservative inference ($RI-p = 0.33$), it yields a statistically significant reduction in the propensity-score matched analysis (see Section 8.1). This indication of improved retention shows that binding course deadlines foster persistence rather than inducing exit. Moreover, by forcing early engagement, the policy strengthened the institutional commitment and academic habits necessary for persistence, as outlined by the academic-momentum framework (Attewell et al., 2012).

Table 4: Long-Term Effects of Academic Dismissal Policy: Business Administration

	Credits (1)	GPA (2)	Dropout (3)	On-Time Grad. (4)	Grad. ≤ 9 Sem. (5)
DiD Estimate	21.86*	-0.01	-0.05	0.04	0.07*
	(1.67)	(0.01)	(0.01)	(0.01)	(0.01)
RI-t p-value	[0.05]	[0.81]	[0.33]	[0.34]	[0.07]
Bootstrap p-value	[0.08]	[0.58]	[0.25]	[0.22]	[0.08]
Pre-treatment Mean	130.48	2.43	0.26	0.096	0.571
Effect Size (%)	16.8%	-0.5%	-21.2%	41.5%	12.2%
Controls	Yes	Yes	Yes	Yes	Yes
Cohort & Program FE	Yes	Yes	Yes	Yes	Yes
Observations	11,877	11,877	11,877	11,877	10,211

Notes: The table reports DiD estimates for Business Administration, obtained by OLS, for outcomes measured at Semester 7 and for graduation outcomes. All specifications include individual controls (high school GPA, Abitur type, gender, age, first-university-semester indicator, foreign citizenship) as well as cohort and study-program fixed effects. Standard errors clustered at the program level (15 clusters) are shown in parentheses. Wild-cluster bootstrap p-values and RI-t p-values based on randomization inference are reported in brackets. Effect sizes are computed relative to the pre-treatment mean. “On-Time Graduation” denotes completing the degree in Semester 7; “Graduated ≤ 9 Semesters” excludes the 2018 cohort. Significance stars are based on RI-t p-values: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

7.3 Business Administration: Event Study Evidence

Figure 1 presents event study graphs to assess the validity of the identification strategy of the treatment effects in BA. I focus on the outcomes most directly tied to degree progress: credit accumulation (Panels A and B) and graduation (Panels C and D). Full event study regression results for all outcomes appear in Appendix Table A1.

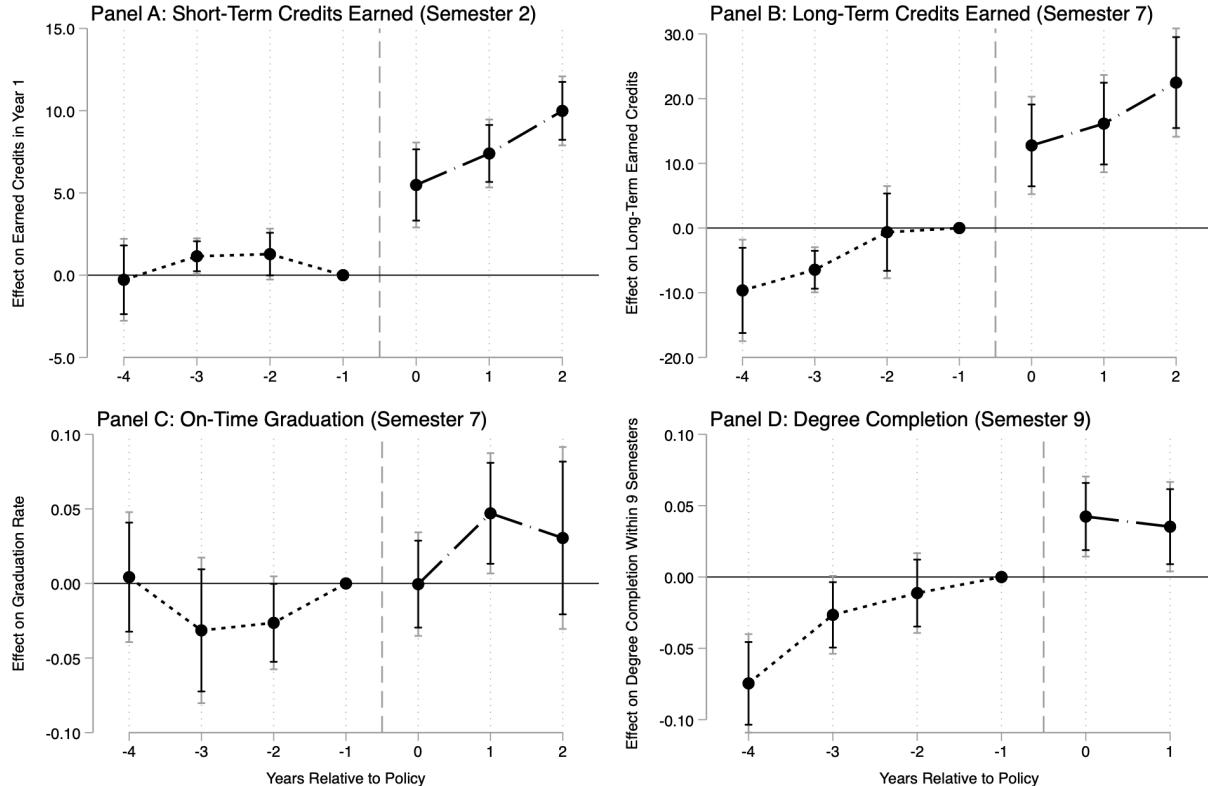


Figure 1: Event Study Graphs: Business Administration

Notes: The figure plots event-study coefficients for Business Administration from specifications estimating dynamic treatment effects relative to the 2015 pre-reform cohort. Panels A and B report effects on credits earned in Semesters 2 and 7; Panels C and D report effects on on-time graduation (Semester 7) and graduation within nine semesters. All models include cohort and program fixed effects and standard student-level controls. Point estimates are shown with 95% confidence intervals based on standard errors clustered at the program level. The vertical dashed line marks the first treated cohort (2016). Full regression output is provided in Appendix Table A1.

Prior to the reform ($k < 0$), coefficients for all outcomes fluctuate around zero with no statistical evidence of systematic drift. Formal validity tests using a wild cluster bootstrap fail to reject the null hypothesis of zero differential pre-trends for semester-two credits ($p = 0.51$), semester-seven credits ($p = 0.31$), on-time graduation ($p = 0.41$), and nine-semester completion ($p = 0.32$). This provides strong statistical support for the parallel pre-treatment trends assumption underlying the DiD design. In the first treated cohort, the event-study coefficients exhibit an upward shift: credit accumulation jumps immediately in 2016 ($K=0$) and remains positive thereafter. Graduation-

related outcomes show analogous discontinuities, with more caution needed for on-time graduation.

7.4 Social Work: Short-Term Effects

The introduction of the relatively lenient and comprehensive deadline in SW generated academic improvements similar to the BA reform, though estimated with less statistical precision. Table 5 reports the short-term effects in SW, measured at Semester 4 when the delayed deadline first binds. Treated students earned 7.8 additional credits by semester four ($SE = 1.86$), roughly a 10% increase over the pre-reform mean. While this point estimate is economically meaningful, it does not reach statistical significance under the conservative randomization inference used for the primary specification ($RI-t p = 0.21$).

This lack of significance reflects the conservative nature of randomization inference in a setting with a single treated cluster. With only 15 programs, the $RI-t$ procedure correctly interprets the substantial between-cluster heterogeneity as noise, resulting in wider confidence intervals. Specifically, precision is constrained by three structural factors: (i) the single treated unit limits variation; (ii) Social Work cohorts are smaller than Business Administration cohorts; and (iii) Social Work students differ demographically from control programs. Consequently, even large point estimates rarely fall in the extreme tails of the placebo distribution, explaining the lack of significance despite favorable magnitudes.

Propensity-score matching directly addresses this compositional imbalance. As shown in Section 8.2, the matched difference-in-differences specification yields a 10.1-credit increase that is statistically significant (bootstrap $p = 0.01$), strongly suggesting that the lack of significance reflects power limitations rather than the absence of a behavioral response. Effects on other short-term outcomes are minimal. Dropout falls by 4.9 percentage points, but again the estimate is statistically indistinguishable from zero ($RI-t p = 0.45$). GPA worsens modestly, although this effect is neither significant nor persistent, as it fully dissipates by semester seven.

Overall, the primary difference-in-differences provides suggestive evidence that the delayed deadline may have shifted behavior in a favorable direction, but the estimates remain too imprecise to draw firm causal conclusions. Stronger statistical evidence emerges from the matched specification (Section 8.2); however, these estimates rely on stronger identifying assumptions and should be interpreted as supportive evidence.

Table 5: Short-Term Effects of Academic Dismissal Policy (Semester 4): *Social Work*

	Credits (1)	GPA (2)	Dropout (3)	Early Dropout (4)
DiD Estimate	7.79 (1.86) [0.21]	0.06 (0.01) [0.15]	-0.04 (0.02) [0.45]	-0.02 (0.01) [0.54]
RI-t p-value				
Bootstrap p-value	[0.20]	[0.38]	[0.32]	[0.46]
Pre-treatment Mean	81.28	2.12	0.20	0.84
Effect Size (%)	9.6%	2.9%	-24.4%	-3.5%
Controls	Yes	Yes	Yes	Yes
Cohort & Program FE	Yes	Yes	Yes	Yes
Observations	11,454	11,454	11,454	5,516

Notes: The table reports DiD estimates for SW, obtained by OLS, for outcomes measured at the end of Semester 2. All models include individual controls (high school GPA, Abitur type, gender, age, first-university-semester indicator, foreign citizenship), cohort fixed effects, and study-program fixed effects. Standard errors clustered at the program level are reported in parentheses. Wild-cluster bootstrap p- and RI-t p-values from randomization inference appear in brackets. Effect sizes are calculated relative to the pre-treatment mean. “Early Dropout” denotes dropout during the first academic year. Significance stars are based on RI-t p-values: *** p<0.01, ** p<0.05, * p<0.10.

7.5 Long-Term Effects: Social Work

Long-term estimates for Social Work are consistent with sustained momentum, though they lack statistical precision in the primary specification (see Table 6). By Semester 7, treated students had accumulated an estimated 12.4 additional credits relative to the control group (SE = 3.23). This point estimate represents a 9% gain over the pre-reform mean, suggesting that students maintained the proportional advantage gained in the first two years rather than falling back to their pre-reform pace. However, under conservative randomization inference, this estimate does not reach statistical significance (RI-t p = 0.23).

This pattern extends to degree completion. The point estimate for on-time graduation indicates an increase of 4.6 percentage points (Column 4), representing a proportional rise of roughly 42% relative to the low pre-reform baseline. As with the short-term results, this estimate is directionally favorable but statistically imprecise (RI-t p = 0.29), reflecting the same power limitations and single-cluster design constraints discussed in Section 7.4.

Despite the lack of significance in the primary specification, four pieces of evidence support the interpretation that these estimates reflect genuine behavioral responses rather than noise. First, the event-study evidence (Section 7.6) shows a structural break in trend at the time of the reform¹. Second, the proportional persistence of the credit effect (10% in the short run, 9% in the long run)

is consistent with the theory of academic momentum rather than mean reversion². Third, the matched difference-in-differences specification (Section 8.2), which restricts the comparison to observably similar students, yields a marginally significant 5.7 percentage point increase in on-time graduation ($p = 0.079$) and confirms the significant increase in credits earned ($p = 0.015$). Fourth, the synthetic difference-in-differences analysis (Section 9.2) decisively resolves the statistical ambiguity of the baseline model, yielding highly significant gains in both long-term credits (11.5 credits, $p = 0.002$) and graduation outcomes. Together, these patterns demonstrate that the delayed deadline activated genuine academic effort, with the baseline imprecision reflecting statistical power constraints rather than a null behavioral response.

Table 6: Long-Term Effects of Academic Dismissal Policy: Social Work

	Credits (1)	GPA (2)	Dropout (3)	On-Time Grad. (4)	Graduated ≤ 9 Semesters (5)
DiD Estimate	12.35 (3.23)	0.01 (0.01)	-0.03 (0.01)	0.04 (0.01)	0.01 (0.01)
RI-t p-value	[0.23]	[0.86]	[0.78]	[0.29]	[0.51]
Bootstrap p-value	[0.23]	[0.62]	[0.37]	[0.24]	[0.43]
Pre-treatment Mean	142.64	2.04	0.22	0.11	0.59
Effect Size (%)	8.7%	0.6%	-14.1%	41.7%	3.3%
Controls	Yes	Yes	Yes	Yes	Yes
Cohort & Program FE	Yes	Yes	Yes	Yes	Yes
Observations	11,454	11,454	11,454	11,454	9,851

Notes: The table reports DiD estimates for Social Work, obtained by OLS, for outcomes measured at Semester 7 and for graduation outcomes. All specifications include individual controls (high school GPA, Abitur type, gender, age, first-university-semester indicator, foreign citizenship) as well as cohort and study-program fixed effects. Standard errors clustered at the program level (15 clusters) are shown in parentheses. Wild-cluster bootstrap p-values and RI-t p-values based on randomization inference are reported in brackets. Effect sizes are computed relative to the pre-treatment mean. “On-Time Graduation” denotes completing the degree in Semester 7; “Graduated ≤ 9 Semesters” excludes the 2018 cohort. Significance stars are based on RI-t p-values: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

7.6 Event Study Evidence: Social Work

To assess the validity of the research design for the Social Work program, Figure 2 presents the event-study estimates. Each panel plots treatment effects on the main progression outcomes by cohort relative to the 2015 pre-reform baseline. While pre-treatment coefficients are visually noisier than in Business Administration due to smaller cohort sizes, they exhibit no systematic upward or downward trend. Formal bootstrap joint tests fail to reject the null hypothesis of parallel

trends for all key outcomes: Semester 4 credits ($p = 0.40$), Semester 7 credits ($p = 0.32$), on-time graduation ($p = 0.32$), and Graduation rate within 9 semesters ($p = 0.33$) (see Appendix Table A2). Post-reform dynamics vary by outcome. Estimates for credit accumulation and on-time graduation move consistently in a positive direction following the 2016 implementation. In contrast, graduation within nine semesters (Panel D) shows no clear structural break.

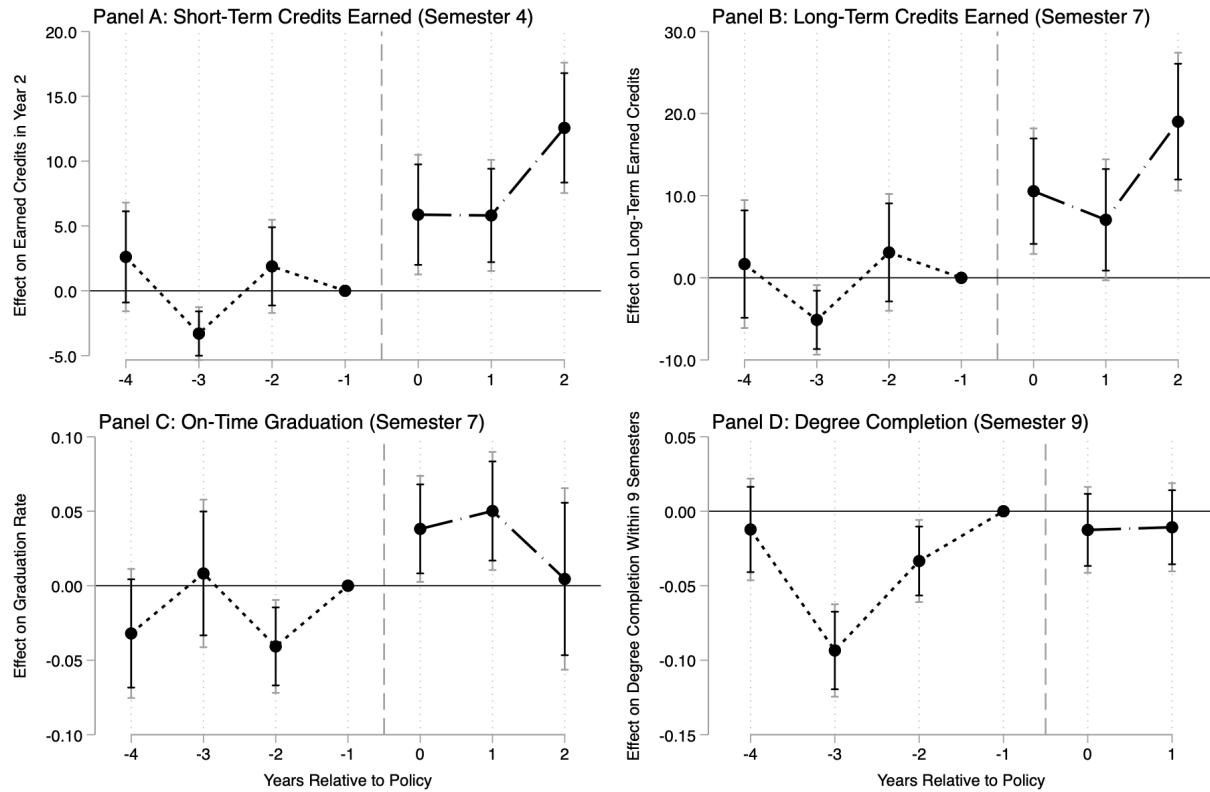


Figure 2: Event Study Graphs: Social Work

Notes: The figure plots event-study coefficients for Social Work from specifications estimating dynamic treatment effects relative to the 2015 pre-reform cohort. Panels A and B report effects on credits earned in Semesters 2 and 7; Panels C and D report effects on on-time graduation (Semester 7) and graduation within nine semesters. All models include cohort and program fixed effects and standard student-level controls. Point estimates are shown with 95% confidence intervals based on standard errors clustered at the program level. The vertical dashed line marks the first treated cohort (2016). Full regression output is provided in Appendix Table A2.

7.6 Heterogeneous Treatment Effects by Student Ability

A central equity concern is whether binding deadlines help all students or disproportionately burden those entering university with weaker academic preparation. Deadlines could, in principle, function as gatekeeping mechanisms that screen out lower-ability students, or they may instead provide external structure that is especially valuable for those who struggle with self-regulation

(Scott-Clayton, 2015). To assess these competing hypotheses, I evaluate whether treatment effects differ across the ability distribution by grouping students into terciles of high-school GPA (High, Medium, Low) and estimating separate difference-in-differences models for each ability group.

Table 7 reports the ability-specific DiD estimates in BA, which serve as the primary specification given their interpretability⁷. These are complemented by a fully interacted triple-difference specification that provides a joint test of equality across groups (Appendix Table A3). Finally, I assess whether pre-treatment ability gaps evolve differently across programs. Event-study graphs (Figure A3) and Wild-bootstrap joint tests fail to reject differential pre-trends for any ability comparison, providing supportive evidence for the identifying assumption (column 4 of Table A3).

The evidence from BA contradicts the gatekeeping hypothesis and supports broad-based effectiveness. Short-term effects (Table 7, Panel A) show sizable and statistically significant credit gains for high- and medium-ability students. High-ability students earn 8.3 additional credits by Semester 2 (bootstrap $p = 0.013$), and medium-ability students gain 6.4 credits (bootstrap $p = 0.024$). Low-ability students exhibit the largest point estimate (11.2 credits), though the estimate is statistically insignificant (bootstrap $p = 0.23$). Across ability groups, GPA remains unaffected, indicating that increases in course load do not dilute academic performance.

Long-term effects (Panel B) display the same broad-based momentum. By Semester 7, high-ability students have accumulated 28.6 additional credits ($p = 0.011$) and experience a 10.4 percentage-point reduction in cumulative dropout ($p = 0.012$). Medium-ability students show similar patterns with 19.6 additional credits ($p = 0.062$) and improved graduation outcomes: on-time graduation increases by 6.7 percentage points ($p = 0.061$), while graduation within nine semesters rises by 7.9 percentage points ($p = 0.091$). Low-ability students again display large positive point estimates across outcomes, but limited bootstrap precision prevents rejecting the null (all $p > 0.20$).

Formal tests for heterogeneity fail to reject the null hypothesis that treatment effects are equal across ability groups (wild bootstrap $p > 0.2$). Furthermore, the consistency of large point estimates across all groups suggests the reform improves outcomes broadly rather than concentrating benefits. These findings indicate that institutional deadlines reduce procrastination frictions across

⁷ Inference for heterogeneous effects relies on the wild cluster bootstrap. While randomization inference (RI-t) is valid for the full sample, subsampling partitions the data into groups where the single treated cluster may become distributionally "atypical" in terms of size or error variance relative to the control distribution. This imbalance violates the exchangeability assumption required for exact inference, potentially leading to severe size distortions (MacKinnon & Webb, 2020). In contrast, the wild cluster bootstrap remains robust to such heterogeneity.

the ability distribution without exacerbating achievement gaps. Even high-ability students respond meaningfully to temporal constraints, suggesting that autonomous scheduling creates inefficiencies even among the well-prepared.

Table 7: Heterogeneous Treatment Effects by Student Ability: Business Administration

<i>Panel A. Short-Term Effects (Semester 2)</i>				
	Credit Points (1)	GPA (2)	Dropout (3)	Early Dropout (4)
High Ability	8.34** (0.97) [0.01]	-0.05 (0.03) [0.41]	-0.04** (0.01) [0.22]	0.01 (0.02) [0.56]
Observations	3,851	3,851	3,851	3,851
Medium Ability	6.42** (0.84) [0.02]	-0.02 (0.02) [0.42]	-0.005 (0.01) [0.75]	0.02 (0.02) [0.51]
Observations	3,881	3,881	3,881	3,881
Low Ability	11.21 (1.51) [0.23]	0.01 (0.02) [0.66]	-0.02 (0.03) [0.50]	0.09 (0.02) [0.41]
Observations	4,127	4,127	4,127	4,127

<i>Panel B. Long-Term Effects (Semester 7)</i>				
	Credit Points	GPA	Dropout	On-Time Graduation
				Graduated \leq 9 Semesters
High Ability	28.57** (3.01) [0.01]	-0.07 (0.02) [0.30]	-0.10** (0.01) [0.01]	0.04 (0.02) [0.31]
Observations	3,851	3,851	3,851	3,851
Medium Ability	19.61* (3.71) [0.06]	-0.00 (0.02) [0.90]	-0.03 (0.02) [0.31]	0.06* (0.01) [0.06]
Observations	3,881	3,881	3,881	3,881
Low Ability	31.51 (5.52) [0.24]	-0.005 (0.02) [0.90]	-0.09 (0.03) [0.36]	0.03 (0.01) [0.29]
Observations	4,127	4,127	4,127	4,127

Notes: This table reports heterogeneous treatment effects estimated separately for high-, medium-, and low-ability students. Each coefficient comes from a distinct difference-in-differences regression estimated within an ability group. Standard errors, clustered at the program level, are reported in parentheses. Wild cluster bootstrap p-values are reported in brackets and serve as the basis for statistical inference. Ability terciles are defined using the distribution of high school GPA in the pre-treatment cohorts***p < 0.01, **p < 0.05, *p < 0.10, based on wild-bootstrap p-values.

Heterogeneity in Social Work—Heterogeneous effects in Social Work (Appendix Tables A4–A5) generally mirror the positive patterns observed in Business Administration, though estimates are less precise. While high-ability students show statistically significant gains in credits and retention, estimates for medium- and low-ability students are directionally positive but statistically insignificant². Formal triple-difference tests fail to reject the null hypothesis of equal treatment effects across ability groups (bootstrap $p > 0.30$). Importantly, no ability group exhibits negative outcomes, supporting the conclusion that the policy did not act as a selective barrier for less-prepared students.

8. Robustness Checks: Propensity-Score Matched Difference-in-Differences

To assess the sensitivity of the estimates to observable compositional differences, I re-estimate the main specifications using a propensity-score-matched difference-in-differences (matched DiD) approach (Heckman et al., 1997). Although DiD with covariates identifies effects under parallel trends, sizeable demographic gaps—especially the pronounced gender imbalance in Social Work—can inflate residual variance and weaken precision in a small-cluster setting⁸. Matching restricts comparisons to observably similar students, improving overlap and reducing noise. Section 9 extends this by using Synthetic DiD to directly optimize pre-treatment outcome trends.

Propensity scores are estimated via a logit model using pre-treatment covariates (high-school GPA, degree type, gender, age, and cohort), and matching is implemented using radius matching with a caliper of 0.10 and replacement. Appendix Figures A3 and A4 document that matching substantially improves covariate balance for both the BA and SW programs. I then re-estimate Equation (1) in the matched sample using matching weights and the same fixed-effects structure as in the primary specification.

8.1 Matched DiD: Business Administration

For BA, matching confirms the robustness of the primary results and sharpens inference on attrition and graduation outcomes, as reported in Table 8. In the short-term (Panel A), treated students accumulated 8.0 additional ECTS credits (bootstrap $p=0.001$) compared to the estimated

⁸ Although DiD with covariates already adjusts for observable differences, linear regression can rely on extrapolation when treated and control groups differ sharply. Matching prevents such extrapolation and improves precision by enforcing overlap, while leaving identification assumptions unchanged.

7.2 in the primary DiD specification. GPA remains statistically unchanged ($p=0.181$), confirming credit gains occurred without quality dilution. While total first-year attrition did not change, the share of eventual dropouts exiting during the first year increased by 9.9 percentage points (bootstrap $p=0.007$). This significant result, which was imprecise in the primary specification, confirms that the deadline accelerates the exit of poorly matched students without increasing overall dropout rates.

Table 8. Matched DiD Results: Business Administration

<i>Panel A. Short-Term Effects (Semester 2)</i>				
	Credit Points (1)	GPA (2)	Dropout (3)	Early Dropout (4)
DiD (ATT)	7.99*** (0.91) [0.000]	-0.02 (0.01) [0.18]	-0.02 (0.01) [0.24]	0.099*** (0.01) [0.007]
Pre-treatment mean	40.27	2.57	0.13	0.52
Effect size (%)	19.86	-0.91	-17.07	18.97
R-squared	0.14	0.15	0.06	0.03
Cohort & Program FE	Yes	Yes	Yes	Yes
Observations	11,842	11,842	11,842	5,754

<i>Panel B. Long-Term Effects (Semester 7)</i>					
	Credit Points (1)	GPA (2)	Dropout (3)	On-Time Graduation (4)	
				Graduated ≤ 9 Semesters (5)	
DiD (ATT)	24.53*** (3.18) [0.000]	-0.019 (0.009) [0.16]	-0.069* (0.01) [0.058]	0.049* (0.01) [0.07]	0.090** (0.01) [0.02]
Pre-treatment mean	130.48	2.43	0.26	0.10	0.57
Effect size (%)	18.80	-0.80	-26.80	51.26	15.87
R-squared	0.10	0.15	0.09	0.09	0.06
Cohort & Program FE	Yes	Yes	Yes	Yes	Yes
Observations	11,842	11,842	11,842	11,842	10,178

Notes: Propensity scores are estimated via logit using high school GPA, degree type, gender, age, and cohort. Matching uses radius matching with caliper 0.10, with replacement and common support. Matched DiD estimates are obtained by weighted OLS with cohort and program fixed effects. standard errors are clustered at the program level and reported in parentheses. Wild cluster bootstrap p-values are reported in brackets.

Long-term effects are similarly robust. By Semester 7, treated students earned 24.5 additional credits (bootstrap $p = 0.001$), aligning closely with the primary estimate. Cumulative dropout falls

by 6.9 percentage points (bootstrap $p=0.058$), moving from statistically insignificant in the main DiD to marginal significance. Finally, the matched specification yields sharper inference for degree completion outcomes: On-time graduation increases by 4.9 percentage points (bootstrap $p=0.078$), and nine-semester graduation by 9 percentage points (bootstrap $p=0.027$). These results reinforce the conclusion that the reform accelerated academic progress and improved degree completion without lowering academic standards. The matched event study (Figure 3) reinforces these findings by achieving tighter pre-trend alignment than the baseline model.

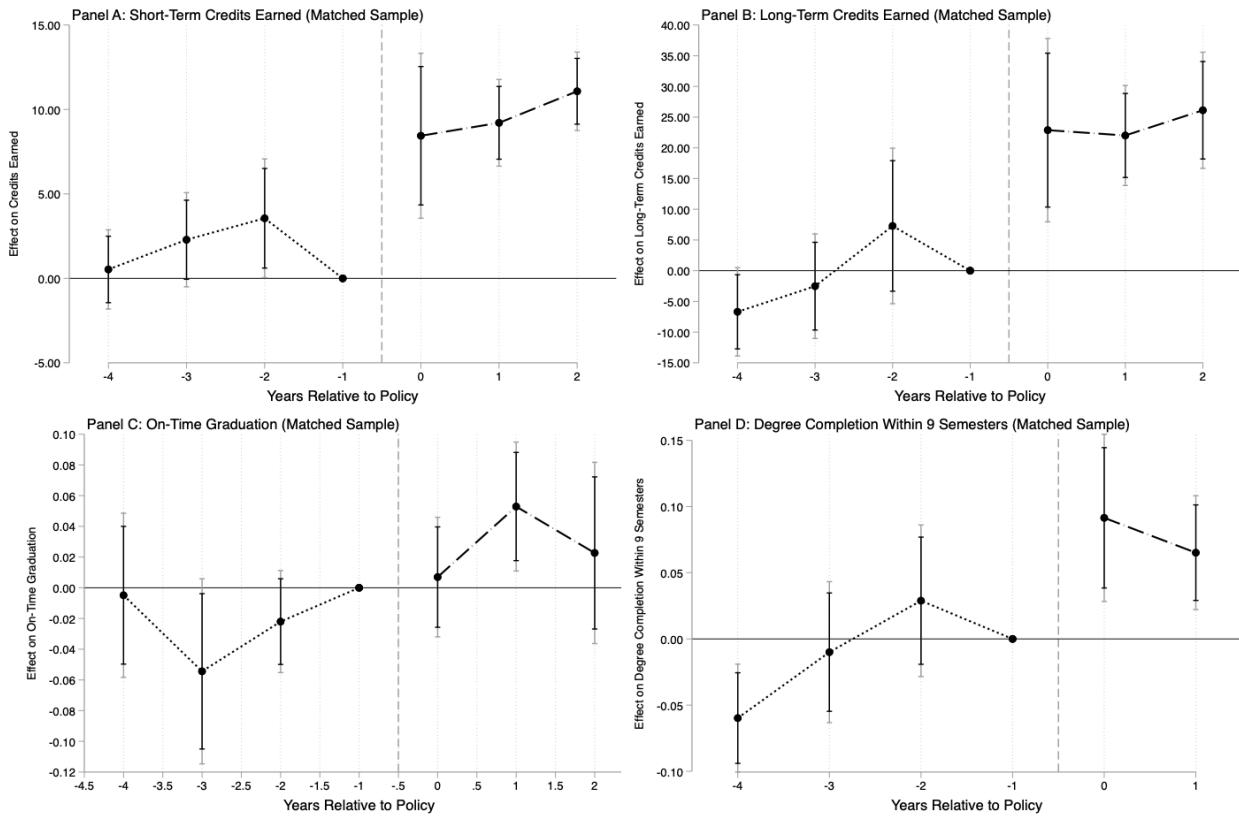


Figure 3: Matched Sample Event Study Graphs: Business Administration *Notes:* This figure plots event-study coefficients for Business Administration using the propensity-score-matched sample, with matching on covariates implemented within cohorts via radius matching. All estimates are relative to the 2015 pre-reform cohort (omitted reference).

8.2 Matched DiD: Social Work

For Social Work, where treated and control programs differ more sharply in the gender composition, the matched DiD approach delivers substantial precision gains while retaining nearly the entire control group. Table 9 reports the matched DiD estimates. In Semester 4, treated students accumulated 10.1 additional credits (bootstrap $p=0.017$), a 12.4 percent increase over the pre-treatment mean. This estimate is both larger and more precisely measured than the primary

estimate of 7.8 credits ($p=0.212$), indicating that balancing on observables substantially improves estimation precision⁹.

One cautionary note emerges: matching detects a small, statistically significant decline in short-term GPA ($p=0.015$), likely reflecting temporary adjustment costs as students accelerated their pace. However, this effect fully dissipates by Semester 7 ($p=0.22$), indicating no lasting penalty on academic quality.

Table 9. Matched DiD Results: Social Work

<i>Panel A. Short-Term Effects (Semester 4)</i>				
	Credit Points (1)	GPA (2)	Dropout (3)	Early Dropout (4)
DiD (ATT)	10.10** (2.08) [0.016]	0.08** (0.01) [0.015]	-0.06 (0.02) [0.12]	-0.02 (0.02) [0.49]
Pre-treatment mean	81.28	2.12	0.20	0.84
Effect size (%)	12.43	4.03	-35.51	-2.38
R-squared	0.11	0.33	0.10	0.02
Cohort & Program FE	Yes	Yes	Yes	Yes
Observations	11,432	11,432	11,432	5,500

<i>Panel B. Long-Term Effects (Semester 7)</i>					
	Credit Points (1)	GPA (2)	Dropout (3)	On-Time Graduation (4)	
				Graduated ≤ 9 Semesters (5)	
DiD (ATT)	17.02** (3.53) [0.015]	0.02 (0.01) [0.22]	-0.05 (0.02) [0.14]	0.057* (0.01) [0.07]	0.050 (0.02) [0.22]
Pre-treatment mean	142.64	2.04	0.22	0.11	0.59
Effect size (%)	11.93	1.43	-26.38	52.98	8.51
R-squared	0.11	0.34	0.10	0.10	0.05
Cohort & Program FE	Yes	Yes	Yes	Yes	Yes
Observations	11,432	11,432	11,432	11,432	9,831

Notes: Propensity scores are estimated via logit using high school GPA, degree type, gender, age, and cohort. Matching uses radius matching with caliper 0.10, with replacement and common support. Matched DiD estimates are obtained by weighted OLS with cohort and program fixed effects. standard errors are clustered at the program level and reported in parentheses. Wild cluster bootstrap p-values are reported in brackets.

⁹ The matched DiD improves comparability by restricting the sample to students with similar propensity scores, effectively reducing heterogeneity-driven noise. This stabilization of the bootstrap distribution allows previously imprecise effects, especially in the Social Work program, to reach standard significance levels. Ultimately, matching resolves the trade-off between conservative inference and estimation precision by minimizing between-cluster variance.

Long-term outcomes show a similar pattern. By Semester 7, treated students have accumulated 17.0 additional credits (bootstrap $p = 0.015$), and on-time graduation increases by 5.7 percentage points (bootstrap $p = 0.079$), whereas these effects were too imprecise to interpret in the primary DiD. The matched results, therefore, clarify that the earlier statistical uncertainty reflects demographic heterogeneity and limited power, not weak behavioral responses. Overall, matching resolves the ambiguity of the baseline estimates and demonstrates that the delayed deadline generated meaningful and sustained academic momentum.

Figure 4 presents the dynamic treatment effects for the primary outcomes of interest—credit accumulation and degree completion—using the propensity-score matched sample. By restricting comparisons to observably similar students, the matching procedure significantly stabilizes the pre-treatment trajectories, which now fluctuate more closely around zero than in the primary event study.

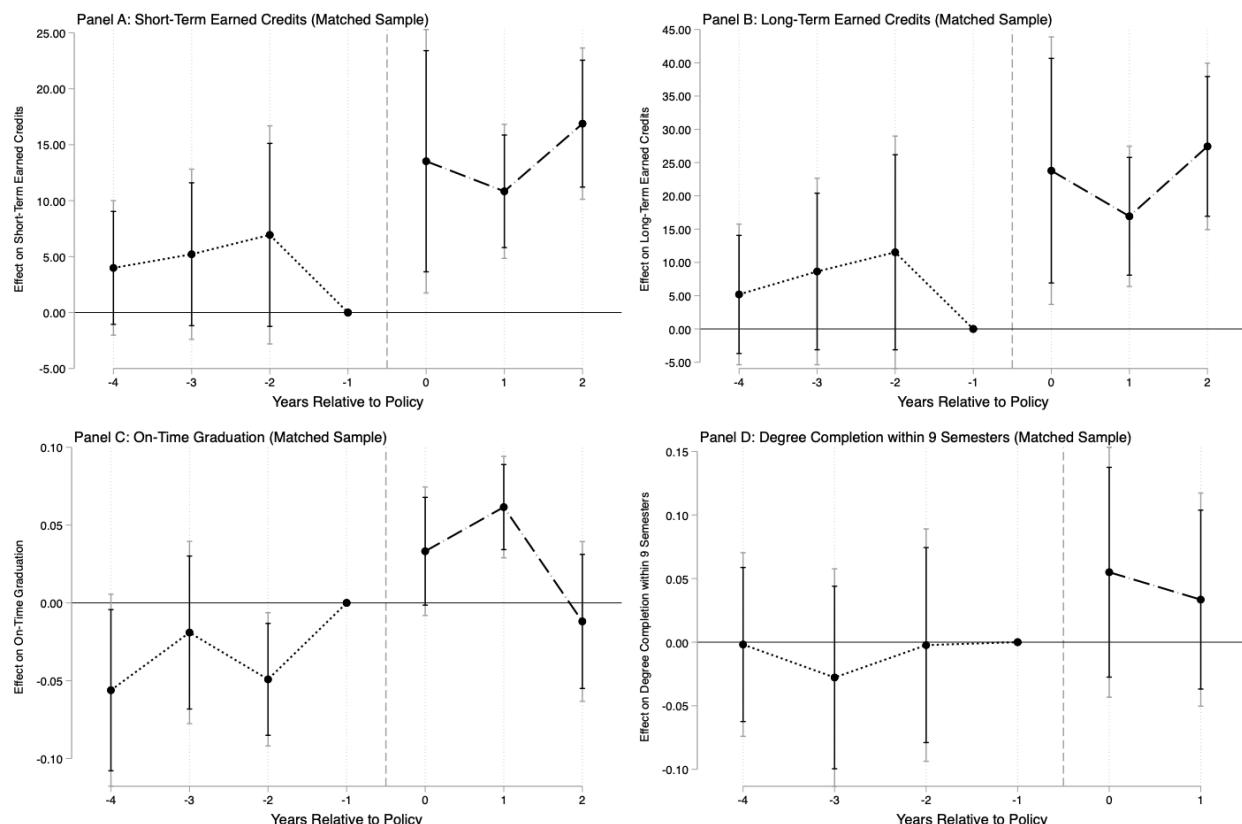


Figure 4: Matched Sample Event Study Graphs: Social Work. *Notes:* This figure plots event-study coefficients for Social Work using The propensity-score-matched sample, with matching on covariates implemented within cohorts via radius matching. All estimates are relative to the 2015 pre-reform cohort (omitted reference).

9. Robustness Checks: Synthetic Difference-in-Differences

To provide an additional layer of robustness, I re-estimate treatment effects using Synthetic Difference-in-Differences (SDID), following Arkhangelsky et al. (2021). SDID constructs a data-driven weighted combination of control units and time periods that optimally matches the treated unit's pre-treatment trajectory, directly minimizing pre-treatment fit discrepancies. Unlike standard DiD, which imposes uniform weights, or matched DiD, which enforces covariate balance, SDID optimizes weights to achieve the closest possible pre-treatment fit on the outcome itself, providing a complementary identification strategy. All specifications utilize optimized covariate adjustment for student background (Abitur, age, and gender), with inference conducted via bootstrap resampling of control units. By constructing a more precise synthetic counterfactual, SDID is particularly well suited to settings in which treated and control programs may differ in levels or trends, even when formal pre-trend tests do not reject parallel trends.

9.1 Synthetic DiD: Business Administration

Table 10, Panel A, reports the short-term SDID estimates for the BA program. In the short term (Semester 2), treated students accumulated 6.1 additional credits ($SE = 1.04$, bootstrap $p < 0.001$), a 15% increase relative to the pre-reform mean. This estimate is broadly consistent with the primary DiD (7.2 credits) and matched-DiD (8.0 credits) results, though modestly more conservative, reflecting the optimized reweighting of control units. The reform also altered the timing of exits: early dropout increases by 10.5 percentage points (bootstrap $p < 0.001$), while total dropout remains unchanged, confirming a significant effect on accelerated sorting rather than increased attrition.

Long-term effects (panel B) confirm that these early gains persist. Treated students earned 12.5 additional credits ($SE = 3.75$, bootstrap $p = 0.001$), smaller than the primary DiD estimate (21.9 credits) but precisely estimated. These credit gains translated into faster degree completion: on-time graduation rises by 4.5 percentage points (bootstrap $p < 0.001$), and completion within nine semesters increases by 5.6 percentage points (bootstrap $p = 0.033$). Crucially, both the matched DiD and the synthetic DiD specifications yield robust and statistically significant effects on graduation outcomes, resolving the imprecision observed in the primary framework and providing consistent evidence that the reform accelerated degree completion.

Figure 5 illustrates these dynamics. Panels A and B show that the SDID-weighted synthetic control closely tracks the treated program throughout the pre-reform period, with a sharp divergence at treatment onset. Panels C and D display graduation outcomes, where the treated cohort consistently outperforms its synthetic counterpart.

Table 10. Synthetic Difference-in-Differences Estimates: Business Administration

<i>Panel A. Short-Term Effects (Semester 2)</i>				
	Credit Points (1)	GPA (2)	Dropout (3)	Early Dropout (4)
Synth. DID	6.058*** (1.040)	-0.048* (0.029)	-0.002 (0.024)	0.105*** (0.015)
P-value (bootstrap)	0.000	0.098	0.946	0.000
Effect Size (%)	15.04%	-1.87%	-1.24%	20.05%
Pre-Mean (Treated)	40.270	2.572	0.131	0.522
Controls	Yes	Yes	Yes	Yes
Effective Controls	12.73	12.54	12.72	12.88
Student Observations	11,860	11,860	11,860	5,763
Program-Cohort Obs.	98	98	98	98

<i>Panel B. Long-Term Effects (Semester 7)</i>					
	Credit Points (1)	GPA (2)	Dropout (3)	On-Time Graduation (4)	
				Graduated ≤ 9 Semesters (5)	
Synth. DID	12.525*** (3.754)	0.001 (0.032)	-0.021 (0.024)	0.045*** (0.006)	0.056** (0.026)
P-value (bootstrap)	0.001	0.965	0.382	0.000	0.033
Effect Size (%)	9.60%	0.06%	-7.93%	47.15%	9.83%
Pre-Mean (Treated)	130.479	2.429	0.260	0.096	0.571
Controls	Yes	Yes	Yes	Yes	Yes
Effective Controls	12.68	12.53	12.49	12.42	11.27
Student Observations	11,860	11,860	11,860	11,860	10,195
Program-Cohort Obs.	98	98	98	98	84

Notes: This table reports Synthetic Difference-in-Differences (SDID) estimates following Arkhangelsky et al. (2021). Bootstrap standard errors (resampling control units) are reported in parentheses. All specifications additionally adjust for student-level covariates (high-school GPA, degree type, gender, and age). *Effective Controls* is computed as 1/HHI, where HHI is the Herfindahl index of synthetic control weights. *Effect Size (%)* is computed relative to the treated group's pre-treatment mean. Significance stars based on bootstrap p-values: *** p<0.01, ** p<0.05, * p<0.10.

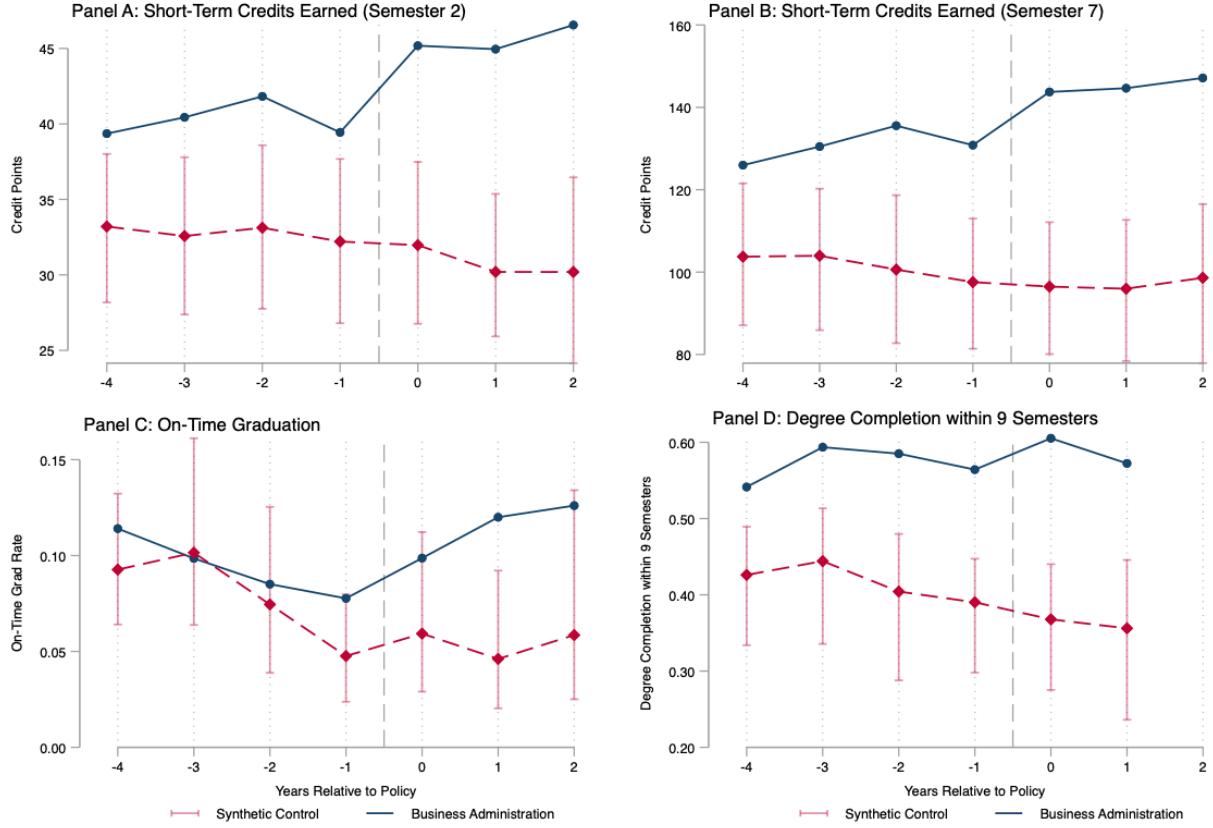


Figure 5: Synthetic Difference-in-Differences Trends in Business Administration. Notes: This figure plots the observed trajectories for the Business Administration program (solid blue) and the synthetic counterfactual (dashed red) using the Arkhangelsky et al. (2021) estimator with optimized unit and time weights. Panels A and B report cumulative credit accumulation at semesters 2 and 7, while Panels C and D display on-time graduation and degree completion within nine semesters.

9.2 Synthetic DiD: Social Work

The SDID estimates for the SW program resolve the statistical imprecision of the primary DiD specifications, delivering more precise estimates that corroborate the matched-DiD findings. In the short run (see Table 11 panel A), treated students accumulated 8.6 additional credits ($p < 0.001$), a 10.6% increase relative to the synthetic counterfactual. These gains persist into the long run, with an additional 11.5 credits by Semester 7 ($p = 0.002$), mirroring the momentum observed in the BA program.

Most importantly, SDID clarifies the reform's effects on degree completion and retention—outcomes that were previously obscured by unweighted trends. On-time graduation increases by 3.5 percentage points ($p = 0.003$), and completion within nine semesters rises by 7.1 percentage points ($p = 0.010$), both of which were statistically insignificant in the primary RI-*t* analysis ($p \geq 0.29$). In the long run (Panel B), there are no detectable effects on GPA or overall dropout,

indicating that the reform increased academic effort and pace without compromising performance or altering attrition rate. By reducing statistical noise and strengthening precision, the SDID results confirm and reinforce the matched-DiD evidence that the SW reform generated meaningful and sustained academic momentum despite its more lenient design.

Table 11. Synthetic Difference-in-Differences Estimates: Social Work

<i>Panel A. Short-Term Effects (Semester 2)</i>					
	Credit Points (1)	GPA (2)	Dropout (3)	Early Dropout (4)	
Synth. DID	8.648*** (2.398)	0.044 (0.028)	-0.044** (0.018)	-0.048* (0.027)	
P-value (bootstrap)	0.000	0.122	0.018	0.076	
Effect Size (%)	10.64%	2.07%	-22.13%	-5.69%	
Pre-Mean (Treated)	81.279	2.116	0.197	0.840	
Controls	Yes	Yes	Yes	Yes	
Effective Controls	12.68	12.17	12.82	12.32	
Student Observations	11,437	11,437	11,437	5,500	
Program-Cohort Obs.	98	98	98	98	
<i>Panel B. Long-Term Effects (Semester 7)</i>					
	Credit Points (1)	GPA (2)	Dropout (3)	On-Time Graduation (4)	Graduated ≤ 9 Semesters (5)
Synth. DID	11.454*** (3.704)	0.005 (0.023)	-0.011 (0.022)	0.035*** (0.012)	0.071*** (0.028)
P-value (bootstrap)	0.002	0.810	0.619	0.003	0.010
Effect Size (%)	8.03%	0.27%	-5.03%	32.46%	12.02%
Pre-Mean (Treated)	142.643	2.038	0.218	0.108	0.593
Controls	Yes	Yes	Yes	Yes	Yes
Effective Controls	12.84	12.56	12.90	12.44	12.28
Student Observations	11,437	11,437	11,437	11,437	9,834
Program-Cohort Obs.	98	98	98	98	84

Notes: This table reports Synthetic Difference-in-Differences (SDID) estimates following Arkhangelsky et al. (2021). Bootstrap standard errors (resampling control units) are reported in parentheses. All specifications additionally adjust for student-level covariates (high-school GPA, degree type, gender, and age). *Effective Controls* is computed as 1/HHI, where HHI is the Herfindahl index of synthetic control weights. *Effect Size (%)* is computed relative to the treated group's pre-treatment mean. Significance stars based on bootstrap p-values: *** p<0.01, ** p<0.05, * p<0.10.

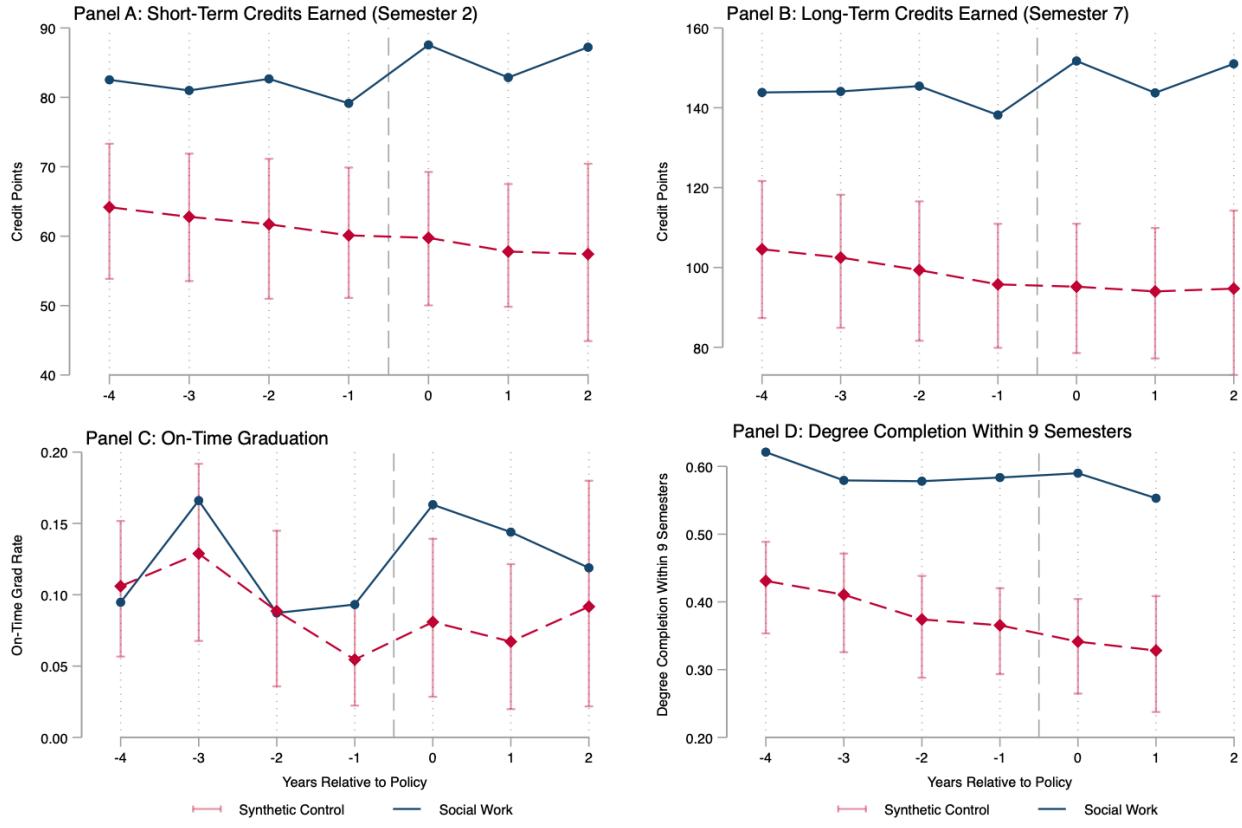


Figure 6: Synthetic Difference-in-Differences Trends in Social Work. Notes: This figure plots the observed trajectories for the Social Work program (solid blue) and the synthetic counterfactual (dashed red) using the Arkhangelsky et al. (2021) estimator with optimized unit and time weights. Panels A and B report cumulative credit accumulation at semesters 2 and 7, while Panels C and D display on-time graduation and degree completion within nine semesters.

10. Discussion and Conclusion

This study demonstrates that mandatory course-completion deadlines substantially improve student outcomes in Germany's tuition-free, high-autonomy higher education system. Exploiting policy reforms where two programs introduced binding deadlines in 2016, I find deadlines increased early credit accumulation by 10–18 percent, long-term accumulation by 9–17 percent, reduced cumulative dropout by roughly 6 percentage points, and raised timely graduation by 5–9 percentage points. These gains emerged without lowering GPA and without increasing overall attrition, indicating a genuine expansion of effort rather than quality dilution or negative selection. Both an early, targeted, high-stakes design (BA) and a later, comprehensive, moderate-stakes design (SW) generated meaningful improvements, though the former produced sharper and more precisely estimated effects. The convergence of results across unweighted, matched, and synthetic

specifications confirms that the observed improvements reflect a systematic behavioral response rather than idiosyncratic trend drift.

Mechanisms: Commitment and Momentum—The results align closely with behavioral models of present bias and procrastination (O'Donoghue and Rabin, 1999, 2001). In environments with weak short-run consequences for delay and high autonomy, students have incentives to postpone costly effort. This dynamic is pervasive: approximately 80% to 95% of college students procrastinate, with nearly half doing so chronically in ways that impair performance (Steel, 2007). Binding deadlines raise the immediate cost of inaction and therefore function as externally imposed commitment devices, analogous to penalty-based contracts shown to increase worker effort in labor markets (Ariely and Wertenbroch 2002; Kaur, Kremer, and Mullainathan 2015). Given the widespread prevalence of procrastination, these interventions hold promise for benefiting a broad student population, not just those with acute self-regulation challenges.

Consistent with this mechanism, recent evidence from a large-scale field experiment demonstrates that offering students commitment devices directly increases the number of credits they earn (Himmler, Jäckle, and Weinschenk 2019). Similarly, Patterson (2018) demonstrates that external pacing tools significantly improve persistence in unstructured learning environments by preventing students from falling behind. In the present setting, stable GPAs and unchanged overall dropout rates reinforce this interpretation: students responded to deadlines not by reallocating effort or strategically exiting, but by expanding total effort in a way consistent with commitment-driven behavioral change.

Crucially, the early acceleration triggered self-reinforcing academic momentum. Short-run credit gains persisted proportionally into later semesters, translating into higher completion rates. Three complementary channels likely operate: (i) accumulation of academic human capital and study habits that compound over time (Stinebrickner and Stinebrickner, 2008, 2014); (ii) strengthened social and institutional integration through regular engagement (Tinto, 2012; Astin, 1999); and (iii) psychological commitment via sunk costs and enhanced self-efficacy (Arkes and Blumer, 1985; Bandura, 1997). The persistence of proportional effects rules out simple front-loading and provides some of the first quasi-experimental evidence that institutions can actively generate momentum rather than merely observe its correlates (Attewell et al., 2012; Belfield et al., 2016).

Policy Design and Institutional Context—These findings directly relate to the debate on academic dismissal policies. Some evidence from the Dutch system shows that strict credit thresholds mainly function through selection, causing dropout to occur earlier without improving graduation rates (Arnold 2015; Cornelisz et al. 2020). In contrast, our results reveal clear incentive effects: credits and graduation rates increase, GPA stays stable, and overall dropout does not rise. Such patterns are inconsistent with cream-skimming and align with genuine behavioral responses. The only selective effect is that students who would have dropped out anyway leave earlier, indicating better information rather than the exclusion of potentially successful students. This distinction has important welfare implications. Recent evidence suggests that students near dismissal thresholds who persist to graduation earn substantial returns (Ost, Pan, and Webber, 2018), indicating that policies facilitating completion rather than exit generate meaningful human capital gains even for marginal students.

The contrast with the Dutch results likely reflects differences in policy design. The German deadlines function as a targeted and fully recoverable form of academic dismissal: they impose meaningful progression requirements that strengthen student effort, yet remain mild enough to avoid the discouragement and attrition effects often seen under harsher, system-wide dismissal regimes (Lindo et al., 2010).

To understand which specific design features drive these incentives, it is necessary to examine the interaction of timing, scope, and flexibility. A comparison of the BA and SW reforms reveals distinct approaches: BA imposed early, targeted deadlines with tight constraints, while SW employed broader and later requirements with more leniency in implementation. Interestingly, these features appear to counterbalance one another; the breadth of the later deadline was diluted by its flexibility, whereas the narrow scope of the early deadline was amplified by its immediacy. The fact that both reforms produced similar improvements suggests that, within highly flexible systems, early intervention can effectively substitute for comprehensiveness by shaping behavior during the critical period of study habit formation.

Equity and Economic Magnitude—A frequent concern is that performance standards disproportionately harm less-prepared students (Scott-Clayton, 2015). The heterogeneity analysis provides reassuring evidence to the contrary: treatment effects were large and positive across the entire high-school-GPA distribution, with point estimates often largest (though noisier) for low-ability students. Deadlines appear to function as a universal pacing device that reduces behavioral

frictions for all students. This uniformity is consistent with U.S. findings from “structured pathway” reforms in U.S. community colleges, where clearer milestones and reduced choice complexity help the least-prepared students stay on track (Scrivener et al., 2015).

A central advantage of the deadline reforms is that they generate substantial academic gains at virtually zero marginal cost. Unlike resource-intensive interventions such as personalized coaching (Bettinger and Baker 2014), summer bridge programs (Barnett et al. 2012), or large-scale financial aid expansions (Goldrick-Rab et al. 2016), adjusting progression rules requires only administrative changes. In a publicly funded system, such structural nudges can yield unusually high returns.

A back-of-the-envelope calculation illustrates the magnitude. Consider the Business Administration program, where the reform increased graduation within nine semesters by 7 percentage points. Public expenditure on tertiary education in Germany averages roughly €11,700 per student-year (OECD 2023), and the median starting salary for business graduates is about €45,000 (Heming et al., 2020). If the marginal graduates enter the labor market just one semester earlier, the combined public savings and additional earnings amount to approximately €28,000 per student. For a typical cohort of 250 students, this implies total social benefits of around €500,000 per cohort, generated entirely by a low-cost structural change that reduces excess time-to-degree.

Limitations and Future Research—Several limitations qualify the interpretation of these findings and point toward productive avenues for future work. First, the analysis relies on reforms within a single university. While the institutional setting is representative of Germany’s flexible, low-tuition system, the single-treated-unit design inherently limits external validity. Second, the comparison group consists predominantly of STEM programs. Although event-study tests support the parallel trends assumption, I cannot rule out unmeasured, time-varying shocks specific to STEM fields, such as labor market shifts or funding changes. Future research should exploit variation in deadline policies within the same academic field across multiple institutions to further isolate the treatment effect from sectoral trends.

Third, identification relies on administrative outcomes and therefore cannot disentangle the underlying behavioral mechanisms. The patterns are consistent with momentum and commitment-device channels, but direct evidence on study time or effort is unavailable. Furthermore, the net impact on well-being is ambiguous: deadlines may increase short-term pressure while simultaneously relieving the chronic stress of procrastination. Future research should combine

administrative records with survey data, digital traces, or well-being metrics to definitively isolate these pathways and evaluate the full welfare trade-offs. Finally, linking education records to social security earnings data is essential to determine whether the reform affects long-run labor market success (e.g., through habit formation) or simply shifts the timing of entry without altering wage trajectories.

In conclusion, this study demonstrates that simple structural policies can substantially improve outcomes in flexible academic settings where weak short-run incentives often lead to drift. By replacing open-ended timelines with clear, recoverable milestones, universities provide necessary commitment devices that activate early momentum. Crucially, these gains do not compromise equity or rigor: deadlines improved progress across the entire ability distribution without increasing attrition. More broadly, the results highlight the central role of institutional design in shaping educational achievement. As higher education faces growing demands for efficiency, these findings suggest that calibrated deadlines offer a scalable, zero-cost tool to support student success without sacrificing access.

References

- Akerlof, G. A. (1991). Procrastination and obedience. *The American Economic Review*, 81(2), 1–19.
- Ariely, D., & Wertenbroch, K. (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science*, 13(3), 219–224.
- Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes*, 35(1), 124–140.
- Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, 111(12), 4088–4118.
- Arnold, I. J. (2015). The effectiveness of academic dismissal policies in Dutch university education: An empirical investigation. *Economics of Education Review*, 40(6), 1068–1084.
- Arnold, I. J. (2023). A high bar may benefit weak students. *Higher Education*, 86(5), 1027–1047.
- Astin, A. W. (1999). Student involvement: A developmental theory for higher education. *Journal of College Student Development*, (pp. 251–262). Routledge..
- Attewell, P., Heil, S., & Reisel, L. (2012). What is academic momentum? And does it matter? *Educational Evaluation and Policy Analysis*, 34(1), 27–44.
- Attewell, P., & Monaghan, D. B. (2016). How many credits should an undergraduate take? *Research in Higher Education*, 57(6), 682–713.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. (Vol. 11). Freeman.
- Barnett, E. A., Bork, R. H., Mayer, A. K., Pretlow, J., Washington, H. D., & Weiss, M. J. (2012). Bridging the Gap: An Impact Study of Eight Developmental Summer Bridge Programs in Texas. *National Center for Postsecondary Research*.
- Bäulke, L., & Dresel, M. (2023). Higher-education course characteristics relate to academic procrastination: A multivariate two-level analysis. *Educational Psychology*, 43(4), 263–283.
- Belfield, C., Jenkins, D., & Lahr, H. (2016). *Momentum: The academic and economic value of a strong start* (CCRC Working Paper No. 88). Community College Research Center, Columbia University.
- Bettinger, E. P., & Baker, R. (2014). The effects of student coaching: An evaluation of a randomized experiment in student advising. *Educational Evaluation and Policy Analysis*, 36(1), 3–19.

- Bisin, A., & Hyndman, K. (2020). Present-bias, procrastination and deadlines in a field experiment. *Games and Economic Behavior*, 119, 339–357.
- Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology*, 67(3), 366.
- Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3), 414–427.
- Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317–372.
- Cornelisz, I., van der Velden, R., de Wolf, I., & van Klaveren, C. (2020). The consequences of academic dismissal for academic success. *Studies in Higher Education*, 45(11), 2175–2189.
- Goldrick-Rab, S., Kelchen, R., Harris, D. N., & Benson, J. (2016). Reducing income inequality in educational attainment: Experimental evidence on the impact of financial aid on college completion. *American Journal of Sociology*, 121(6), 1762–1817.
- Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, 64(4), 605–654.
- Heming, J., Stanski, C., & Zimmermann, T. (2020). *Gehaltsreport für Absolventen 2020/21*. StepStone.
https://www.stepstone.de/content/de/de/5/projects/campus/download/StepStone_2020_Gehaltsreport_fuer_Absolventen.pdf
- Heublein, U., Ebert, J., Hutzsch, C., Isleib, S., König, R., Richter, J., & Woisch, A. (2017). Zwischen studienerwartungen und studienwirklichkeit. Ursachen des Studienabbruchs, beruflicher Verbleib der Studienabbrecherinnen und Studienabbrecher und Entwicklung der Studienabbruchquote an deutschen Hochschulen, 1.
- Himmller, O., Jäckle, R., & Weinschenk, P. (2019). Soft commitments, reminders, and academic performance. *American Economic Journal: Applied Economics*, 11(2), 114–142.
- Kaur, S., Kremer, M., & Mullainathan, S. (2015). Self-control at work. *Journal of Political Economy*, 123(6), 1227–1291.
- Kőszegi, B. (2010). Utility from anticipation and personal equilibrium. *Economic Theory*, 44(3), 415–444.
- Lavecchia, A. M., Liu, H., & Oreopoulos, P. (2016). Behavioral economics of education: Progress and possibilities. In *Handbook of the Economics of Education* (Vol. 5, pp. 1-74).

- Lindo, J. M., Sanders, N. J., & Oreopoulos, P. (2010). Ability, gender, and performance standards: Evidence from academic probation. *American Economic Journal: Applied Economics*, 2(2), 95–117
- MacKinnon, J. G., & Webb, M. D. (2018). The wild bootstrap for few (treated) clusters. *The Econometrics Journal*, 21(2), 114–135.
- MacKinnon, J. G., & Webb, M. D. (2020). Randomization inference for difference-in-differences with few treated clusters. *Journal of Econometrics*, 218(2), 435–450.
- O'Donoghue, T., & Rabin, M. (1999). Doing it now or later. *The American Economic Review*, 89(1), 103–124.
- O'Donoghue, T., & Rabin, M. (2001). Choice and procrastination. *The Quarterly Journal of Economics*, 116(1), 121–160.
- OECD. (2023). *Education at a Glance 2023: OECD Indicators*. OECD Publishing.
- Ost, B., Pan, W., & Webber, D. (2018). The returns to college persistence for marginal students: Regression discontinuity evidence from university dismissal policies. *Journal of Labor Economics*, 36(3), 779-805.
- Ostermaier, A. (2018). Incentives for students: effects of certificates and deadlines on student performance. *Journal of Business Economics*, 88(1), 65-96.
- Patterson, R. W. (2018). Can behavioral tools improve online student outcomes? Experimental evidence from a massive open online course. *Journal of Economic Behavior & Organization*, 153, 293–321
- Schmidt, H. G., Baars, G. J., Hermus, P., van der Molen, H. T., Arnold, I. J., & Smeets, G. (2022). Changes in examination practices reduce procrastination in university students. *European Journal of Higher Education*, 12(1), 56-71.
- Scott-Clayton, J. (2015). The shapeless river: Does a lack of structure inhibit students' progress at community colleges?. In *Decision making for student success* (pp. 102-123). Routledge.
- Scrivener, S., Weiss, M. J., Ratledge, A., Rudd, T., Sommo, C., & Fresques, H. (2015). Doubling graduation rates: Three-year effects of CUNY's Accelerated Study in Associate Programs (ASAP) for developmental education students. New York, MDRC
- Sneyers, E., & De Witte, K. (2017). The effect of an academic dismissal policy on dropout, graduation rates and student satisfaction: Evidence from the Netherlands. *Studies in Higher Education*, 42(2), 354–389.
- Statistisches Bundesamt. (2024). *Statistischer Bericht - Statistik der Studierenden - Wintersemester 2023/2024*. Wiesbaden. <https://www.destatis.de>

- Steel, P. (2007). The nature of procrastination: a meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological bulletin*, 133(1), 65.
- Stinebrickner, R. and Stinebrickner, T.R. (2008). The causal effect of studying on academic performance. *B.E. Journal of Economic Analysis and Policy* 8(1): 1–53
- Stinebrickner, R., & Stinebrickner, T. R. (2014). Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model. *Journal of Labor Economics*, 32(3), 601–644.
- Tinto, V. (2012). *Leaving college: Rethinking the causes and cures of student attrition*. University of Chicago Press.
- Zamir, E., Lewinsohn-Zamir, D., & Ritov, I. (2017). It's now or never! using deadlines as nudges. *Law & Social Inquiry*, 42(3), 769-803.

APPENDIX

Table A1. Event-Study Difference-in-Differences Estimates: Business Administration

	Short-term Effects				Long-term Effects				
	Credits Sem 2 (1)	GPA Sem 2 (2)	Dropout Sem 2 (3)	Early Dropout (4)	Credits Sem 7 (5)	GPA Sem 7 (6)	Dropout Sem 7 (7)	On-Time Graduation (8)	Graduated ≤ 9 Semesters (9)
t-4 Coefficient	-0.2818 (1.2708)	-0.0223 (0.0172)	0.0465 (0.0278)	-0.1098** (0.0449)	-9.6424** (4.0040)	-0.0052 (0.0131)	0.0436* (0.0229)	0.0042 (0.0222)	-0.0745*** (0.0176)
t-3 Coefficient	1.1493* (0.5536)	-0.0207 (0.0217)	0.0203** (0.0087)	-0.1336*** (0.0217)	-6.4534*** (1.7826)	0.020. (0.0229)	0.0120 (0.0119)	-0.0315 (0.0249)	-0.0265* (0.0139)
t-2 Coefficient	1.2801 (0.7889)	-0.1063*** (0.0154)	-0.0017 (0.0233)	-0.1253*** (0.0239)	-0.6416. (3.6340)	-0.0695*** (0.0159)	0.0063. (0.0199)	-0.0264 (0.0159)	-0.0112 (0.0142)
t0 Coefficient	5.4828*** (1.3174)	-0.0860** (0.0345)	0.0407 (0.0262)	-0.0551** (0.0222)	12.7664*** (3.8458)	-0.0420 (0.0284)	-0.0261. (0.0235)	-0.0004 (0.0177)	0.0423*** (0.0143)
t+1 Coefficient	7.3980*** (1.0533)	-0.0642* (0.0320)	0.0039 (0.0242)	-0.1054*** (0.0255)	16.1375*** (3.8405)	-0.0296 (0.0257)	-0.0185. (0.0196)	0.0470** (0.0206)	0.0352** (0.0159)
t+2 Coefficient	9.9875*** (1.0706)	-0.0194 (0.0283)	-0.0212 (0.0239)	0.0168. (0.0283)	22.4770*** (4.2744)	-0.0093 (0.0231)	-0.0677** (0.0232)	0.0305 (0.0311)	N/A N/A
Pre-treatment Mean	40.2702	2.5724	0.1305	0.5217	130.4793	2.4285	0.2597	0.0957	0.5705
R-squared	0.3084	0.3175	0.0940	0.0344	0.2449	0.3474	0.1961	0.1537	0.2146
Observations	11,877	11,877	11,877	5,779	11,877	11,877	11,877	11,877	10,211
<i>Panel B: Parallel Trends Test</i>									
Joint Test P-val.	0.5185	0.0641	0.4665	0.2643	0.3183	0.0961	0.5556	0.4114	0.3232

Notes: This table reports event study estimates using the full trend group. The policy was implemented for the 2016 cohort; the 2015 cohort is used as the reference group, so all coefficients are relative to this baseline. “Graduated within 9 Semesters” uses cohorts from 2012 to 2017; the 2018 cohort is excluded because students could only be observed through their 8th semester. t-2 corresponds to two cohorts prior to treatment, t0 is the treatment cohort, t+1 is the 2017 cohort, etc.. Panel A reports the estimated coefficients with standard errors in parentheses. Panel B provides p-values for joint tests of pre-treatment coefficients (t-4 through t-1), which assess the parallel trends assumption; “Pass” indicates that the null hypothesis of equal pre-trends cannot be rejected at conventional levels. This assumption is formally tested using a joint hypothesis test of pre-treatment coefficients. Parallel Trends Test Specification: $H_0: \beta_{\{t-4\}} = \beta_{\{t-3\}} = \beta_{\{t-2\}} = 0$. Parallel trends testing employs a wild cluster bootstrap with Webb weights (999 replications) to test the joint null hypothesis. The bootstrap method is essential given the small number of study program clusters, as standard asymptotic tests can provide unreliable p-values when cluster counts are low. For most outcomes, the parallel trends assumption holds, while GPA outcomes show marginal evidence of differential pre-trends.

Table A2. Event-Study Difference-in-Differences Estimates: SW

	Short-term Effects				Long-term Effects				
	Credits Sem 2 (1)	GPA Sem 2 (2)	Dropout Sem 2 (3)	Early Dropout (4)	Credits Sem 7 (5)	GPA Sem 7 (6)	Dropout Sem 7 (7)	On-Time Graduation (8)	Graduated ≤ 9 Semesters (9)
t-4 Coefficient	2.6135. (2.1387)	-0.1137*** (0.0142)	0.0072. (0.0256)	-0.0331. (0.0254)	1.6716 (3.9746)	-0.0725*** (0.0128)	0.0101 (0.0229)	-0.0320 (0.0221)	0.1020 (0.1042)
t-3 Coefficient	-3.2954*** (1.0399)	-0.0568** (0.0208)	0.0283. (0.0176)	-0.0374. (0.0225)	-5.1197** (2.1612)	-0.0228 (0.0215)	0.0270* (0.0141)	0.0083 (0.0253)	0.2137** (0.0718)
t-2 Coefficient	1.8837. (1.8345)	-0.1037*** (0.0156)	-0.0232. (0.0264)	-0.1000*** (0.0236)	3.0926 (3.6299)	-0.0748*** (0.0160)	-0.0332 (0.0199)	-0.0408** (0.0159)	0.2848*** (0.0807)
t0 Coefficient	5.8738** (2.3556)	0.0323. (0.0279)	-0.0399. (0.0247)	-0.1403*** (0.0166)	10.5449** (3.9035)	0.0102 (0.0297)	-0.0151 (0.0242)	0.0381* (0.0182)	-0.1696* (0.0895)
t+1 Coefficient	5.8112** (2.1888)	-0.0146. (0.0265)	-0.0153. (0.0237)	-0.0391. (0.0224)	7.0606* (3.7596)	-0.0288 (0.0267)	-0.0037 (0.0196)	0.0502** (0.0202)	-0.2849** (0.1050)
t+2 Coefficient	12.5624*** (2.5656)	-0.0285. (0.0226)	-0.0794*** (0.0260)	-0.0229. (0.0224)	19.0120*** (4.2860)	-0.0672*** (0.0225)	-0.0667** (0.0233)	0.0045 (0.0311)	N/A N/A
Pre-treatment Mean	81.2789	2.1161	0.1967	0.8400	142.6429	2.0376	0.2177	0.1077	8.4023
R-squared	0.2730	0.4162	0.1762	0.0363	0.2502	0.4392	0.2023	0.1544	0.1892
Observations	11454	11454	11454	5516	11454	11454	11454	11454	4718
<i>Panel B: Parallel Trends Test</i>									
Joint Test P-val.	0.4064	0.1532	0.5005	0.4384	0.3213	0.4685	0.2893	0.3213	0.3303

Notes: This table reports event study estimates using the full trend group. The policy was implemented for the 2016 cohort; the 2015 cohort is used as the reference group, so all coefficients are relative to this baseline. “Graduated within 9 Semesters” uses cohorts from 2012 to 2017; the 2018 cohort is excluded because students could only be observed through their 8th semester. t-2 corresponds to two cohorts prior to treatment, t0 is the treatment cohort, t+1 is the 2017 cohort, etc.. Panel A reports the estimated coefficients with standard errors in parentheses. Panel B provides p-values for joint tests of pre-treatment coefficients (t-4 through t-1), which assess the parallel trends assumption; “Pass” indicates that the null hypothesis of equal pre-trends cannot be rejected at conventional levels. This assumption is formally tested using a joint hypothesis test of pre-treatment coefficients. Parallel Trends Test Specification: $H_0: \beta_{-4} = \beta_{-3} = \beta_{-2} = 0$. Parallel trends testing employs a wild cluster bootstrap with Webb weights (999 replications) to test the joint null hypothesis. The bootstrap method is essential given the small number of study program clusters, as standard asymptotic tests can provide unreliable p-values when cluster counts are low. For most outcomes, the parallel trends assumption holds, while GPA outcomes show marginal evidence of differential pre-trends.

Table A3. Heterogeneous Treatment Effects: Triple-Difference Estimates (BA)

Outcome	High Ability (1)	Medium Ability (2)	Low Ability (3)	Joint HTE P-val (4)	Diff. Pre-trend H – M (5)	Diff. Pre-trend L – M (6)
<i>Panel A. Short-Term Effects (Semester 2)</i>						
Credit Points	9.003** (1.038) [0.010]	6.372** (0.826) [0.032]	10.815 (1.562) [0.265]	0.267 N=11,877	0.286	0.364
GPA	0.057 (0.039) [0.390]	0.027 (0.026) [0.462]	0.002 (0.025) [0.947]	0.562 N=11,877	0.665	0.632
Dropout	-0.055 (0.017) [0.189]	-0.005 (0.015) [0.729]	-0.024 (0.031) [0.553]	0.245 N=11,877	0.143	0.352
Early Dropout	0.114 (0.033) [0.234]	0.054 (0.016) [0.274]	0.088 (0.02) [0.396]	0.353 N=5,779	0.218	0.367
<i>Panel B. Long-Term Effects (Semester 7)</i>						
Credit Points	31.083*** (3.216) [0.007]	20.148* (3.727) [0.067]	29.32 (5.815) [0.302]	0.262 N=11,877	0.457	0.48
GPA	-0.077 (0.025) [0.284]	-0.001 (0.025) [0.959]	-0.009 (0.025) [0.755]	0.398 N=11,877	0.46	0.469
Dropout	-0.117** (0.015) [0.011]	-0.043 (0.024) [0.306]	-0.082 (0.035) [0.416]	0.231 N=11,877	0.271	0.427
On-Time Graduation	0.049 (0.026) [0.294]	0.071** (0.012) [0.047]	0.029 (0.011) [0.380]	0.38 N=11,877	0.14	0.144
Graduated ≤ 9 Semesters	0.109* (0.017) [0.053]	0.083* (0.017) [0.096]	0.117 (0.023) [0.287]	0.531 N=10,211	0.293	0.459

Notes: Columns 1–3 report the treatment effects for each ability group estimated using separate DiD regressions. Standard errors clustered at the program level are reported in parentheses. Wild cluster bootstrap p-values are reported in brackets. "Joint HTE" reports the bootstrap p-value for the null hypothesis that treatment effects are equal across all three ability groups. "Diff. Pre-trend" reports the p-value for a joint test of parallel pre-treatment trends in ability gaps relative to the medium-ability group. Early Dropout is conditional on eventual dropout (N=5,779). Graduated ≤ 9 Semesters excludes the 2018 cohort. Significance stars based on bootstrap p-values: *** p<0.01, ** p<0.05, * p<0.10.

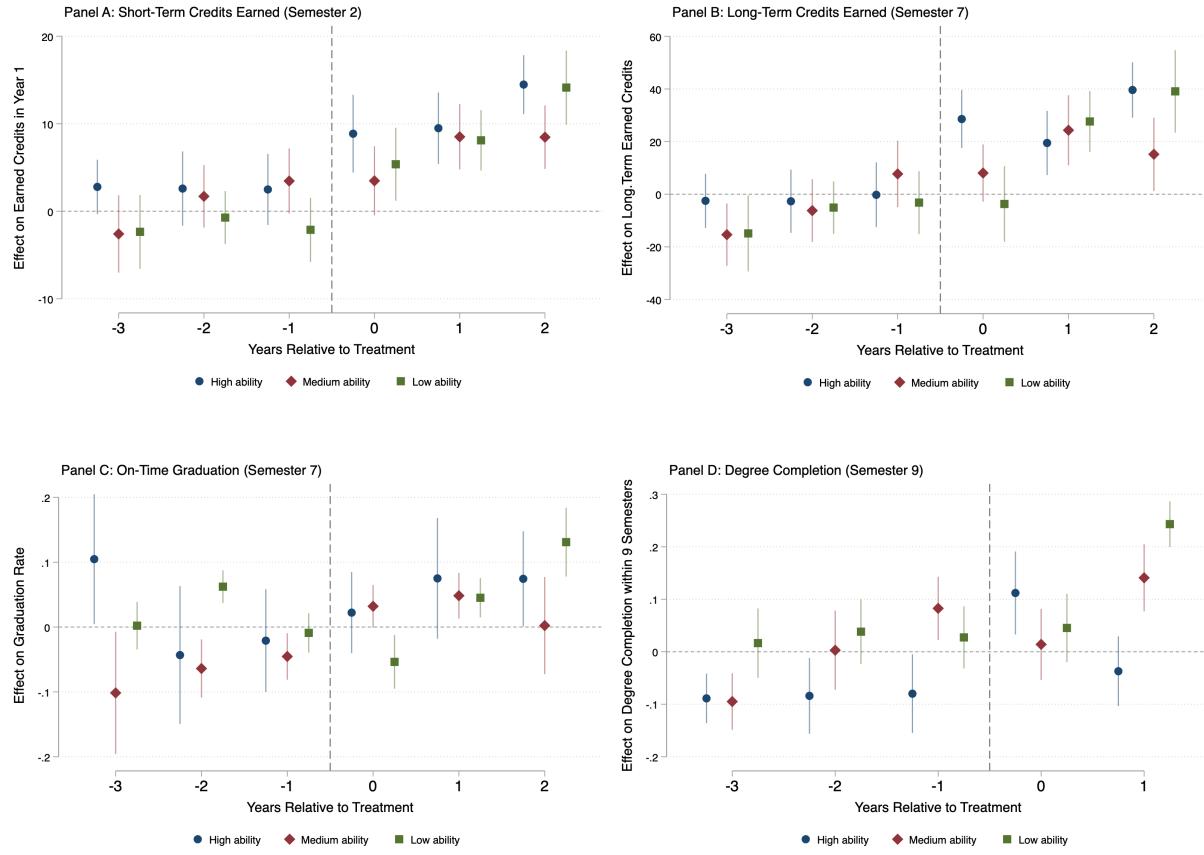


Figure A1. Event-Study Estimates by Ability Group: Business Administration

Notes: This figure reports event-study estimates of the effects of mandatory course-completion deadlines by student ability group. Panels display outcomes for short-term credits earned (Semester 2), long-term credits earned (Semester 7), on-time graduation (Semester 7), and degree completion (Semester 9). Ability groups are defined by terciles of high school GPA in the pre-treatment cohorts. Points represent estimated coefficients relative to the 2015 cohort ($t = 0$), with vertical bars indicating 95 percent confidence intervals based on wild cluster bootstrap standard errors clustered at the program level. Negative values indicate outcomes below the reference cohort mean.

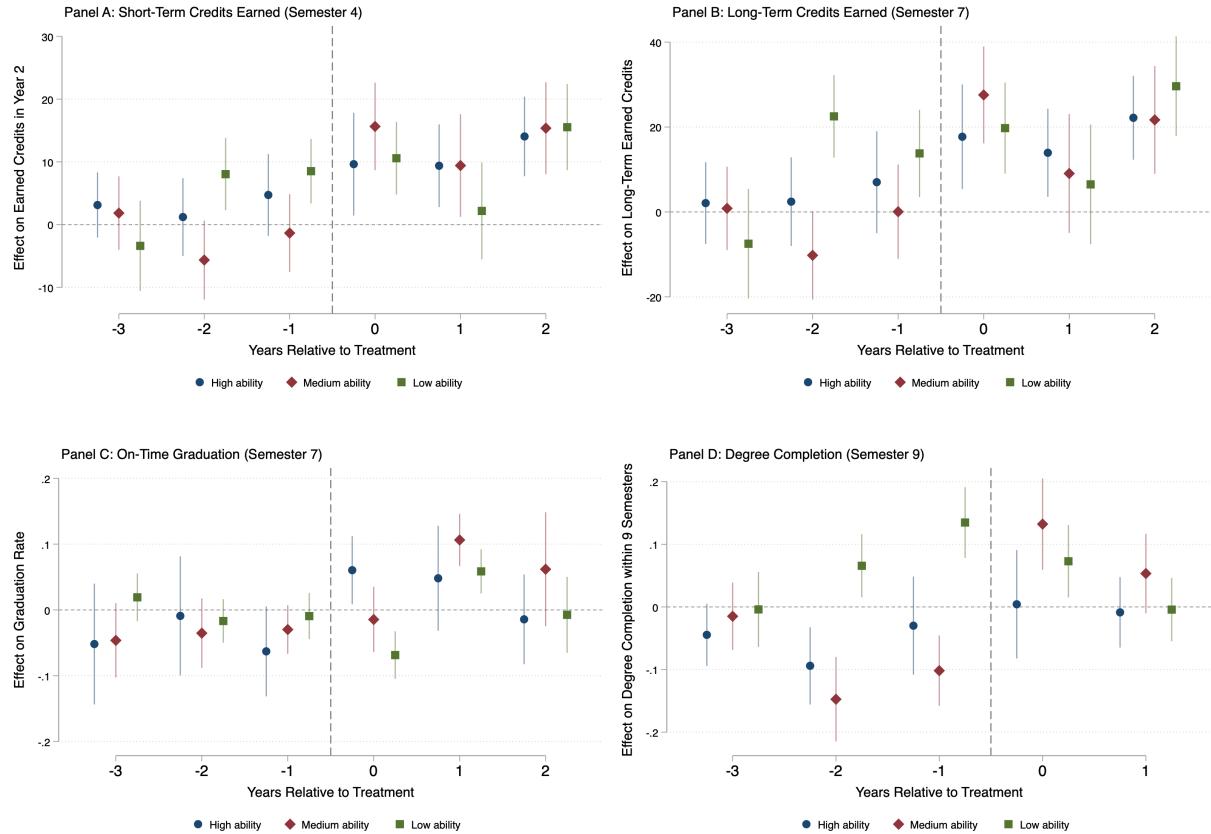


Figure A2. Event-Study Estimates by Ability Group: Social Work

Notes: This figure reports event-study estimates of the effects of mandatory course-completion deadlines by student ability group. Panels display outcomes for short-term credits earned (Semester 2), long-term credits earned (Semester 7), on-time graduation (Semester 7), and degree completion (Semester 9). Ability groups are defined by terciles of high school GPA in the pre-treatment cohorts. Points represent estimated coefficients relative to the 2015 cohort ($t = 0$), with vertical bars indicating 95 percent confidence intervals based on wild cluster bootstrap standard errors clustered at the program level. Negative values indicate outcomes below the reference cohort mean.

Table A4: Heterogeneous Treatment Effects by Student Ability: SW

<i>Panel A. Short-Term Effects (Semester 2)</i>				
	Credit Points (1)	GPA (2)	Dropout (3)	Early Dropout (4)
High Ability	8.918** (2.008) [0.114]	0.023 (0.027) [0.499]	-0.077** (0.016) [0.076]	-0.062 (0.030) [0.278]
Observations	4,080	4,080	4,080	1,106
Medium Ability	14.503** (2.593) [0.104]	0.011 (0.028) [0.746]	-0.085 (0.029) [0.275]	-0.167*** (0.016) [0.149]
Observations	3,306	3,306	3,306	1,556
Low Ability	6.578 (2.681) [0.520]	0.134 (0.025) [0.348]	-0.062 (0.033) [0.487]	-0.007 (0.023) [0.760]
Observations	4,050	4,050	4,050	2,838

<i>Panel B. Long-Term Effects (Semester 7)</i>					
	Credit Points (5)	GPA (6)	Dropout (7)	On-Time Graduation (8)	Graduated \leq 9 Semesters (9)
High Ability	15.266** (3.403) [0.115]	-0.021 (0.027) [0.536]	-0.065** (0.015) [0.105]	0.060 (0.025) [0.227]	0.038 (0.018) [0.304]
Observations	4,080	4,080	4,080	4,080	3,417
Medium Ability	20.960** (4.685) [0.152]	-0.066 (0.030) [0.265]	-0.062 (0.029) [0.298]	0.090*** (0.018) [0.165]	0.140*** (0.021) [0.245]
Observations	3,306	3,306	3,306	3,306	2,843
Low Ability	13.580 (4.948) [0.492]	0.087 (0.024) [0.374]	-0.038 (0.031) [0.536]	0.011 (0.011) [0.523]	-0.021 (0.017) [0.414]
Observations	4,050	4,050	4,050	4,050	3,573

Notes: This table reports heterogeneous treatment effects estimated separately for high-, medium-, and low-ability students. Each coefficient comes from a distinct difference-in-differences regression estimated within an ability group. Standard errors, clustered at the program level, are reported in parentheses. Wild cluster bootstrap p-values (Rademacher weights, 999 replications) are reported in brackets and serve as the basis for statistical inference. Ability terciles are defined using the distribution of high school GPA in the pre-treatment cohorts. "On-Time Graduation" indicates completion within six semesters; "Graduated \leq 9 Semesters" indicates completion within nine semesters. Observations refer to the estimation sample for each ability group and outcome horizon. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$, based on wild-bootstrap p-values.

Table A5. Heterogeneous Treatment Effects: Triple-Difference Estimates (SW)

	High Ability (1)	Medium Ability (2)	Low Ability (3)	Joint HTE P-val (4)	Diff. Pre-trend H – M (5)	Diff. Pre-trend L – M (6)
Credit Points	10.318* (2.003) [0.033]	14.095 (2.704) [0.132]	7.997 (3.156) [0.576]	0.367 N: 11,454	0.521	0.401
GPA	0.011 (0.027) [0.741]	-0.000 (0.027) [0.995]	0.146 (0.029) [0.387]	0.798 N: 11,454	0.096	0.649
Dropout	-0.088* (0.016) [0.020]	-0.082 (0.028) [0.271]	-0.075 (0.034) [0.524]	0.862 N: 11,454	0.496	0.479
Early Dropout	-0.072 (0.032) [0.257]	-0.154 (0.018) [0.356]	-0.005 (0.024) [0.849]	0.483 N: 5,516	0.133	0.448
<i>Panel B. Long-Term Effects (Semester 7)</i>						
Credit Points	17.542* (3.277) [0.038]	20.479 (4.818) [0.155]	16.173 (5.540) [0.562]	0.574 N: 11,454	0.556	0.427
GPA	-0.034 (0.026) [0.377]	-0.073 (0.029) [0.276]	0.096 (0.028) [0.387]	0.440 N: 11,454	0.165	0.725
Dropout	-0.077* (0.015) [0.035]	-0.062 (0.029) [0.282]	-0.059 (0.032) [0.562]	0.660 N: 11,454	0.563	0.413
On-Time Graduation	0.067 (0.026) [0.201]	0.092 (0.015) [0.209]	0.001 (0.012) [0.953]	0.403 N: 11,454	0.568	0.486
Graduated \leq 9 Semesters	0.049 (0.017) [0.184]	0.121 (0.022) [0.233]	-0.013 (0.020) [0.567]	0.247 N: 9,851	0.551	0.197

Notes: Columns 1–3 report the treatment effects for each ability group estimated using separate DiD regressions. Standard errors clustered at the program level are reported in parentheses. Wild cluster bootstrap p-values are reported in brackets. "Joint HTE" reports the bootstrap p-value for the null hypothesis that treatment effects are equal across all three ability groups. "Diff. Pre-trend" reports the p-value for a joint test of parallel pre-treatment trends in ability gaps relative to the medium-ability group; values above 0.10 indicate no evidence of differential pre-trends. Early Dropout is conditional on eventual dropout (N=5,779). Graduated \leq 9 Semesters excludes the 2018 cohort. Significance stars based on bootstrap p-values: *** p<0.01, ** p<0.05, * p<0.10.

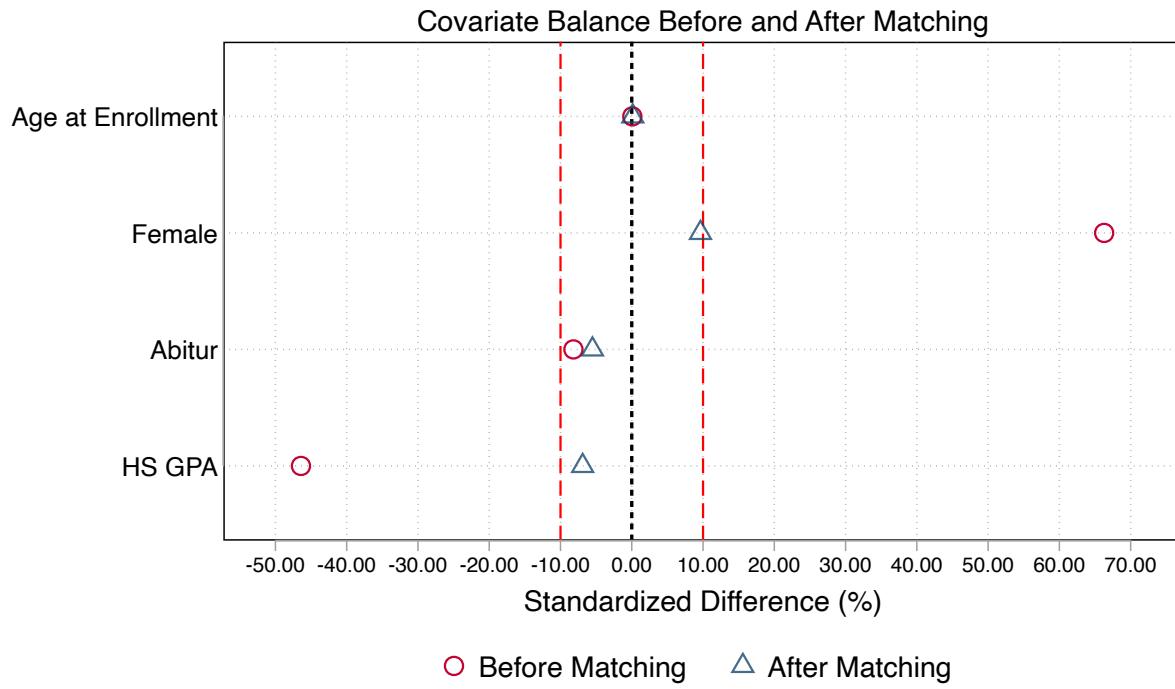


Figure A3. Standardized Covariate Differences Before and After Propensity Score Matching:
Business Administration

Note: This figure displays standardized mean differences for selected covariates before and after propensity score matching. Circles represent unmatched (before matching) differences, and triangles represent matched (after matching) differences. Dashed vertical lines indicate conventional balance thresholds ($\pm 10\%$). Improved balance after matching is indicated by estimates closer to zero.

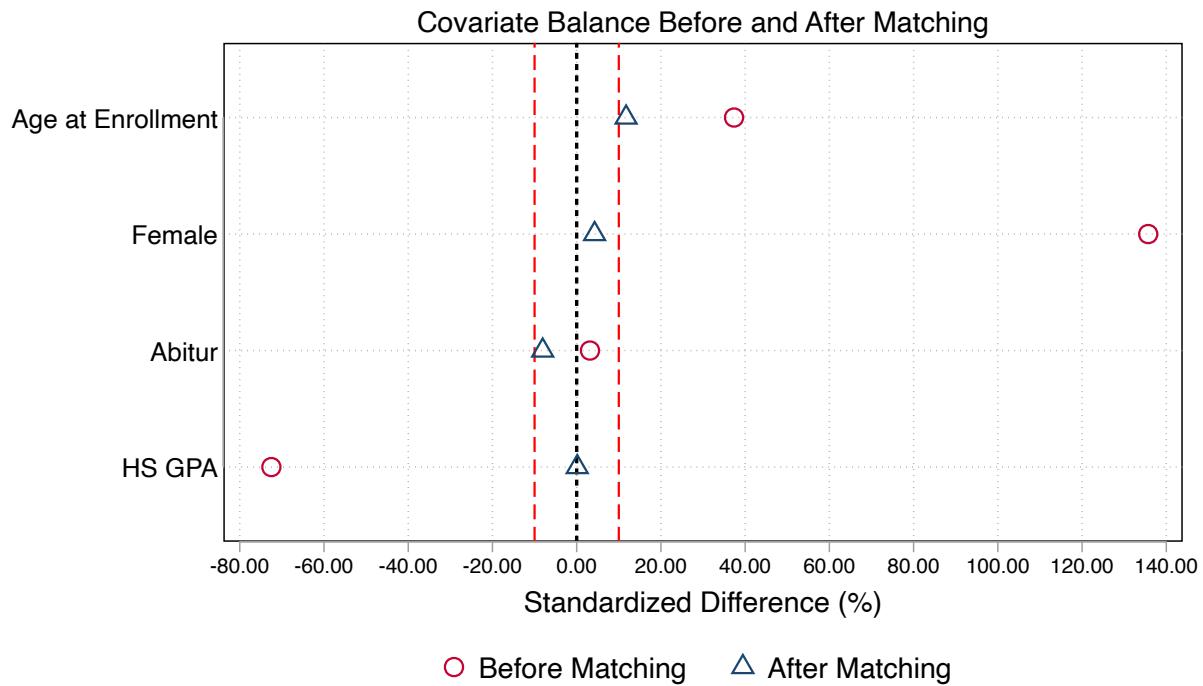


Figure A4. Standardized Covariate Differences Before and After Propensity Score Matching:
SW

Note: This figure displays standardized mean differences for selected covariates before and after propensity score matching. Circles represent unmatched (before matching) differences, and triangles represent matched (after matching) differences. Dashed vertical lines indicate conventional balance thresholds ($\pm 10\%$). Improved balance after matching is indicated by estimates closer to zero.

Table A6. Summary of Regulatory Stability and Non-Deadline Changes

Panel A. Business Administration: Regulatory Adjustments		
Category	Description of Change	Nature and Relevance
Deadline Policy	Introduction of binding deadlines: Modules G1 (“General BWL”) and G2 (“Math”) must be passed by the end of Semester 1.	Primary treatment. The only binding progression constraint introduced.
Module Sizing	Variable module sizes (2–8 ECTS) standardized to a uniform 5-ECTS grid.	Administrative. Facilitates mobility and scheduling; total content volume unchanged.
Renaming	e.g., Einführung in die BWL → Allgemeine BWL; Materialwirtschaft → Logistik.	Cosmetic. Modernized titles; content unchanged.
Sequencing	Two modules (Macroeconomics, Business English) moved from Year 2 to Year 1.	Neutral. Balances workload; curriculum unchanged.
Specializations	Specialization requirement adjusted from 3×14 ECTS to 3×12 ECTS.	Technical. Credits redistributed; total degree remains 210 ECTS.

Panel B. Social Work: Regulatory Adjustments		
Category	Description of Change	Nature and Relevance
Deadline Policy	Introduction of binding deadline: All 60 first-year credits must be passed by the end of Semester 4.	Primary treatment. The only binding progression constraint introduced.
Renaming	e.g., Theorie/Geschichte → Einführung in die Wissenschaft; Theorien der SA → Wissenschaft der SA.	Cosmetic. Updated titles; content unchanged.
Thesis Registration	Registration deadline moved to “one month after start of 9th semester.”	Administrative. Aligns thesis timing with the new rules.
Assessment Forms	Removal of Referat (presentation) as mandatory partial exam in Modules 1.5 and 1.13.	Minor simplification. Reduced assessment complexity; workload unchanged.
Prerequisites	Removal of specific ECTS prerequisites for intermediate modules (e.g., Module 2.5).	Flexibility. Eliminates minor bottlenecks; smoother sequencing.

Notes: Based on the official Study and Examination Regulations (SPO) for Business Administration and Social Work. The 2016 reforms introduced binding course-completion deadlines as the only substantive change affecting student progression incentives. All other adjustments were administrative, cosmetic, or minor technical revisions that did not alter course content, credit requirements, or curricular structure.