IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17400

Another Chance: Number of Exam Retakes and University Students' Outcomes

Massimiliano Bratti
Silvia Granato
Enkelejda Havari

OCTOBER 2024

DISCUSSION PAPER SERIES

IZA DP No. 17400

Another Chance: Number of Exam Retakes and University Students' Outcomes

Massimiliano Bratti
University of Milan, GLO, IZA and LdA

Silvia Granato
JRC

Enkelejda Havari
IESEG School of Management, GLO, IZA and Lille Economics and Management

OCTOBER 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Another Chance: Number of Exam Retakes and University Students' Outcomes*

Exams play a key role in a student's learning process at university, and their organization may affect student performance. A high number of retakes, for instance, could encourage procrastination or reduce effort for each attempt. This article investigates the effects of a policy change at a major Italian university that reduced the number of exam retakes allowed per subject from six to three. Using a difference-in-differences strategy, we find that this policy significantly improved first-year outcomes, including lower dropout rates, higher exam pass rates, and increased credit accumulation. We conduct several robustness checks showing that only a small fraction of these improvements can be attributed to changes in the average quality of students enrolled following the reform. Additionally, the policy contributed to an increase in on-time graduation rates, which was the main objective of the reform, without harming student GPA. This study shows that implementing a cost-effective policy, such as limiting exam retakes, can substantially enhance student progression, reducing age at graduation.

JEL Classification: I21, I23

Keywords: student outcomes, exams, retakes, university, Italy

Corresponding author:
Enkelejda Havari
IESEG School of Management
3 rue de la Digue
59000 Lille
France
E-mail: e.havari@ieseg.fr

* We gratefully acknowledge comments by participants in several departmental seminars and at the European Society for Population Economics (ESPE) 2024 Conference, the 2nd Workshop on Educational Economics and Policy 2024 (WEEP), the 49th Annual Conference the Association for Education Finance & Policy 2024, the European Association of Labour Economists (EALE) 2023 Conference, the 8th LEER Conference on Education Economics 2023, the 10th Italian Congress of Econometrics and Empirical Economics (ICEEE), the 2023 International Workshop on Applied Economics of Education (IWAEE), the 2023 Workshop on "Higher Education and Equality of Opportunities" (Modena), the Counterfactual Methods for Policy Impact Evaluation (COMPIE) 2022 conference, and the 2022 Annual Meeting of the Italian Association of Labour Economics (AIEL). We are grateful to Danilo Cinti, Carmela Tanzillo, and Camilla Valentini from the University of Bologna (Italy) for providing the data used in this study. The usual disclaimers apply.

1 Introduction

Students play a key role in the learning process and are not passive subjects.[1] Consequently, teachers and educational institutions continuously work to motivate and encourage student effort. One key mechanism for doing so is through exams, which provide significant extrinsic motivation. Exams allow students to assess their level of preparation and knowledge while also acting as the primary way they earn their "reward" through grades.

Given the central role of exams in education, it is natural to think that their organization might influence student performance. Yet, it is surprising that exam organization has received far less attention compared to other determinants of educational performance such as, just to mention a few, peer group effects (e.g., Booij et al., 2017; Feld and Zölitz, 2017), college costs (e.g., Garibaldi et al., 2012; Ketel et al., 2016), and financial aid (e.g., Scott-Clayton, 2011; Jones et al., 2022).

The intuition behind this paper is simple. Just as the level of motivation of football players may be very different in a friendly match compared to the final of the Champions League, the level of student motivation and effort may depend on the number of exam retakes that are allowed. If too many retakes are offered, even high-stakes exams—like those taken at university—become low-stakes because students know there will always be "another chance." This abundance of opportunities may lead students to procrastinate or reduce their effort on the early exam attempts. As a result, exam organization may affect student performance.

When it comes to the number of exam retakes, their availability can enhance the likelihood of passing an exam, but it may also lead to a decrease in student effort (Bertola, 2024; Michaelis and Schwanebeck, 2016; Nijenkamp et al., 2016; Kooreman, 2013). Conversely, extended intervals between retakes can lead to knowledge depreciation, which mitigates the adverse impact on student effort. (Nijenkamp et al., 2018).[2] There might be additional learning taking place thanks to retakes (students may study again to resit the exam), especially when exams are high-stakes, and the possibility of retakes is limited (Goodman et al., 2020; Frisancho et al., 2016). Exam schedules that stimulate student attention and com-

[1] Stinebrickner and Stinebrickner (2008), for instance, shows that students' study time is an important determinant of human capital accumulation.

[2] Michaelis and Schwanebeck (2016) show in a model with one main session and one retake that reducing the spacing between the two exams increases student effort and raises the probability of passing the first examination attempt while reducing the probability of passing the second. Simulation results suggest that the overall effect on passing the exam is negative.

mitment (both during and outside lectures) may increase student performance. This may hold true both for intermediate tests (De Paola and Scoppa, 2011; Sulun et al., 2018) and for final exams, for instance, when they are scheduled immediately after the end of lectures (Di Pietro, 2013). There might be calendar effects, and scheduling exams immediately after a relatively long period of holidays (e.g. the Christmas break) may be deleterious to student performance (Di Pietro, 2013). Exam periods too concentrated in time may increase the need for multi-tasking, harming student performance (Schmidt et al., 2022). Overall, there are many forces at play, and the final outcome of a given design of course assessments is theoretically ambiguous.

In this paper, we contribute to this literature by exploring the impact of the number of exam retakes on student outcomes. We use a quasi-experimental design based on a reform implemented by the Faculty of Economics at the University of Bologna (hereafter UniBO),[3] starting from the 2010-11 academic year, which reduced the number of attempts a student could take an exam per subject from six to three.[4]

Italy makes for an interesting case study. Students are granted significant autonomy in organizing their university careers. For instance, they have a high number of exam retakes[5] and can even refuse the grade obtained at exams and retake them. Despite these features of Italian higher education, which should make a student's life easier, Italy has lower continuation rates in higher education, high university dropout (OECD, 2008, Table A4.1),[6] and longer university graduation times (Garibaldi et al., 2012) compared to most OECD countries. As a result, Italy ranks among the lowest OECD countries (second-to-last) in terms of the percentage of the population aged 24-35 with tertiary education, at just 29.2%. This figure is higher only than Mexico's and falls 18 percentage points below the OECD average of 47.4% (2022 data).[7] Thus, Italy provides an ideal setting to examine

[3] UniBO is one of the largest higher education institutions in Italy and the world's oldest university.

[4] Until the academic year 2011-12, the University of Bologna was organized into "Faculties," which grouped degree courses in the same discipline. Starting in 2012-13 a reorganization was carried out, and the University is now constituted of "Schools", which group degree courses in the same broad fields (e.g., the School of Economics, Management and Statistics include degree courses from the pre-existing Faculty of Economics, Faculty of Statistics and Faculty of Management, and some degree courses from the previous Faculty of Political Science). UniBO has campuses and Faculties, not only in the city of Bologna but also in the cities of Cesena, Forlì, Ravenna, and Rimini.

[5] Before the exam reform considered in this paper, a first-semester/first-year exam could be attempted 15 times at the University of Bologna during the legal degree's duration, but this number can further increase for laggard students (the so-called *fuori corso*).

[6] According to OECD data, in 2005 the dropout rate in Italy was around 55% against the OECD average of 31%.

[7] `https://www.oecd.org/en/data/indicators/population-with-tertiary-education.html` (ac-

whether reducing the degrees of freedom for students and aligning exam organization to international standards may improve student outcomes.

UniBO's Faculty of Economics adopted a system that is more in line with international standards and where fewer exam attempts are permitted. For instance, in the United States, retakes are generally not allowed; in the United Kingdom, students typically have only two attempts; and in Sweden, students are usually allowed three attempts per academic year, a structure similar to the one implemented at UniBO. This policy shift not only made the exam organization closer to international standards but also provided a unique opportunity to analyze how reducing the number of exam attempts affects student performance.

Using student administrative data from UniBO, we apply a difference-in-differences (DiD) research design and compare first-year student performance (probability to drop-out, number of credits earned, and number of exams passed) of the cohorts of students enrolled in UniBO degree courses that made the switch to the new system (those taught in the Faculty of Economics) with those that did not introduce this change. Our analysis demonstrates that the reform improved first-year student outcomes. The probability of first-year dropout was reduced by 4.2 percentage points (pp, hereafter), and the number of credits earned and exams passed increased by 11.2 credits and about one exam, respectively.

Since the reform was approved by the University's Council in the spring of 2010 and implemented in September of the same year, it is unlikely that all prospective students intending to enrol were aware of the changes, especially in the first year of the reform. However, rational and well-informed students might have adjusted their enrollment choices accordingly, leading to a "cream-skimming" effect (where higher-quality students are more likely to choose the affected courses). We find that this potential positive selection explains only a small part of the observed improvements —approximately 0.8 additional credits earned and 0.1 more exams passed— with no notable effect on dropout rates. To isolate the effect of the reform from the "cream-skimming," we apply several analytical strategies, including conditional DiD estimation —controlling for student characteristics— and a matching-DiD approach that compares students with similar attributes across cohorts. After accounting for selection, the reform's impact on dropout rates remains of a similar magnitude (-4.5 pp) while its effects on credits earned and exams passed decrease marginally and are 10.1 and 0.9, respectively, in the matching DiD estimates. Additional

cessed 8th October 2024).

robustness checks confirm that our main findings are consistent across different ways of clustering observations and when using doubly-robust DiD estimators (Sant'Anna and Zhao, 2020), and robust to several additional checks to address potential confounding factors.

An exploratory analysis documents that positive effects were not limited to short-run outcomes.[8] Indeed, better performance in the first year was also reflected in longer-term academic outcomes, such as the probability of graduation and of on-time graduation, which increased in the affected degree programs by 5.7 pp and 9.1 pp, respectively. Even more remarkable is the finding that the faster degree completion, driven by the higher likelihood of on-time graduation, did not come at the expense of final graduation marks — a major concern for students.

Our paper demonstrates that, despite facing resistance from student unions, implementing reforms reducing the number of exam attempts can lead to significant improvements in student progression. Yet, while these reforms may improve students' educational outcomes, they may not necessarily raise their satisfaction or well-being — e.g., students might prefer more freedom in allocating study and leisure time over the academic year —, which may cause them to strongly oppose such changes.

The paper proceeds as follows. Section 2 describes the Italian higher education system and the main features of the exam reform that was introduced by the Faculty of Economics of the University of Bologna in the academic year 2010-11. Section 3 summarizes some related literature on the effects of exam schedules, intermediate exams, and exam resits. Data and sample selection are described in Section 4, and the empirical strategy in Section 5. The main results of the empirical analysis, i.e., the effect of the reform on first-year student performance, an event study employed to check the plausibility of the common trend assumption, the "cream skimming" effect, and some robustness checks are commented on in Section 6. This section also discusses some potential mechanisms behind the observed positive effect of the reform and includes an exploratory analysis of long-term academic outcomes such as the probability of graduating and the final graduation mark. Section 7 draws conclusions.

[8] As we discuss later, owing to the characteristics of the reform, some pre-reform cohorts are partially treated when considering long-term academic outcomes.

2 The institutional context

2.1 The Italian higher education system

The Italian higher education system is organized in two cycles: a 3-year first-level degree (‘*Laurea Triennale*’, corresponding to a bachelor’s degree) followed by a 2-year second-level degree (‘*Laurea Specialistica/Magistrale*”, corresponding to a master’s degree). To graduate from a bachelor’s or master’s degree, students must have acquired 180 or 120 credits, respectively. One credit is equivalent to one ECTS (*European Credit Transfer System*) credit, which corresponds to 25 to 30 hours of workload (lecture attendance or study hours). Single-cycle six- or five-year degree programs are envisaged for specific degree courses aimed at training for the exercise of regulated professions and highly qualified activities (including medicine, dentistry, veterinary, pharmacy, architecture, building engineering, law, and primary education sciences). Access to the latter is mostly regulated nationally through an entrance test, while for the rest of the degree courses access is not regulated or single universities are free to autonomously set specific entry requirements.[9] The upper-secondary school curriculum attended does not limit enrolment in tertiary education to specific fields. Italian high school tracks available may be academic — specializing in humanities or science — or vocational, which in turn are articulated in vocational-technical and vocational-professional.[10]

The Italian higher education system offers a high degree of flexibility, enabling students to autonomously organize their study careers. The academic calendar is typically structured in two semesters, with lectures held from September to December and March to May, followed by exam sessions in December-February and June-July, along with an additional session in September. Each exam session provides multiple dates for each subject—typically 2 or 3—offering students numerous opportunities to take exams throughout the academic year. This means that in general students can take an exam for a first-year subject at any time during their enrollment in a degree program.[11] If students do not pass an exam, which requires a minimum grade of 18 out of 30 (with the possibility of

[9] Degree courses with “locally” regulated access are mainly those that involve the use of laboratories, specific IT tools, or internships at external institutions (typically, degrees in the fields of life sciences, engineering, psychology, sport sciences.

[10] Generally, students with the highest academic aptitude tend to enrol in academic tracks, while those with intermediate skills typically choose vocational-technical programs. Students with lower academic performance often pursue vocational-professional tracks.

[11] Some exceptions apply, such as exams that require passing specific “preparatory” subjects first.

receiving a distinction — 30 *cum laude*), or are not satisfied with their grade and refuse it, they have several chances to retake it during their studies.[12] Notably, grading is typically done internally, with assessments carried out by the same faculty members who teach the corresponding courses.

This system imposes a substantial burden on teachers, who must prepare multiple exam versions to prevent cheating, and on universities, which must manage the logistics of organizing numerous exam sessions while ensuring proper supervision. Moreover, they must provide costly services not only to regular students but also to those who are graduating late, which represents a significant cost to society as a whole. In light of the significant costs, the Faculty of Economics at UniBO decided to cut the number of exam retakes from six to three, aiming to streamline the process and alleviate the strain on faculty and administrative resources. After UniBO, other universities have tried to do the same. This topic remains at the forefront of debates in Italian higher education.

2.2 The new exam policy at UniBO

In 2010, the Faculty of Economics of UniBO started a unique policy by reducing the number of exam retakes from 6 to 3, which was proposed and discussed by the Faculty's Council. This initiative was the most significant component of a broader set of actions aimed at improving the efficiency of students' learning.

Table 1 summarizes the main changes introduced by the reform. Up to 2009-10, the exams for the subjects taught in the first semester (fall) of a given academic year $t/t+1$ could be sat in December$_t$-February$_{t+1}$ (3 possible dates), June$_{t+1}$-July$_{t+1}$ (2 possible dates) and September$_{t+1}$ (1 date only). After the reform, from 2010-11 onwards, the schedule for first-semester subjects would change in December-February (2 dates) and September (1 date), with no exam retakes allowed in June-July. Before the reform, the exams for subjects taught in the second semester (spring) were offered in June$_{t+1}$-July$_{t+1}$ (3 possible dates), September$_{t+1}$ (1 date) and in the next academic year, December$_{t+1}$-February$_{t+2}$ (2 dates). After the reform, the exams for these subjects could be taken only in June-July (2 dates) and September (1 date).

The reform came into force for students enrolling in their first year of studies in the academic year 2010-11. Students who started a study career in previous years remained

[12] By refusing a grade, students prevent it from being recorded in their academic records and from contributing to their GPA.

Table 1: Number of exam attempts before and after the 2010-11 reform

Enrolment cohort	Career year	Course Semester	Exam schedule Sessions: Dec.-Feb. #	June-July #	Sept. #
up to 2009-10 (control)	1st	I (fall)	3	2	1
		II (spring)	2	3	1
from 2010-11 (treated)	1st	I (fall)	2 (-1)	0 (-2)	1
		II (spring)	0 (-2)	2 (-1)	1

Note. The figures in parentheses represent the reduction in the number of possible exam attempts that could be sat in each session after the reform. # indicates the number of exam attempts.

under the old exam regime, except for subjects that were taught in previous years. For example, in 2010-11, the 2009-10 cohort could sit up to 6 times an exam of a second-year ("regular" year) subject (old regime), but only up to 3 times an exam of a first-year subject (new regime). For this reason, to have a clean separation between the affected and unaffected cohorts (by the reform), we will focus our main analysis on first-year student performance only.

It is worth mentioning that the exam reform was part of a broader reorganization of the UniBO Faculty of Economics' academic offering, aimed at streamlining it both for a more sustainable and efficient use of resources and for improving students' learning experience. Some of these actions were already adopted in the academic year 2009-10, namely the extension of competency tests at entry to all degree courses, the monitoring of student satisfaction through surveys, and the partial reorganization of the academic calendar. The entry tests, primarily aimed at assessing incoming students' knowledge, placed a stronger emphasis on mathematics. The goal was to identify those who lacked adequate mathematical skills, requiring them to fulfill additional requirements during their studies. This included attending remedial courses or passing the mathematics exams by the end of their first year to qualify for enrollment in their second year.[13] The teaching calendars were also reorganised making them more compact.[14]

[13] In the Faculty of Economics, subjects were generally taught in one semester, except for Mathematics, which was running over two semesters. However, it was divided into different modules: Mathematics I and II, each taught during one semester by different teachers.

[14] In each semester, lectures took place during two sub-periods of 5 weeks (teaching *sub-cycles*), with a week break between them. Typically, bachelor's degree subjects were 8 ECTS (60 lecture hours) subjects taught over the two sub-cycles (10 weeks), while master's degree subjects were 4/5 ECTS (30/40 lecture

The minutes of the Faculty's council and other preparatory documents mention that the stated objective was to increase the regularity of student careers, given the high dropout rates (in 2005-06 and 2006-07, 20% in the second and 25% in the third year) and a high percentage of students graduating with a delay (two-thirds of all students) in the Faculty of Economics of Bologna. Moreover, the members of the council expressed concern with "the significant and rapid reduction of the teaching and research staff and a persistent lack of financial and structural resources."

The change in the organization of exams spurred a hot debate at the regional level, both inside and outside the university. Students opposed the reform, claiming that it was giving too much emphasis on the "speed" of the study career at the cost of reducing the chances for deeper learning. Moreover, students from the Faculty of Economics argued that the reform would put them in a disadvantaged position relative to other Faculties and university students. They also raised the concern that the novelties introduced in the exam organization would harm working students who typically had tighter time constraints and could not attend lectures. These protests eventually led to the introduction of an additional exam attempt starting from the academic year 2014-15 (bringing the total number of exam attempts to 4 for a given subject). However, this extra session could be activated on-demand upon request of interested students.

3 Related literature

This paper is related to three main strands of literature in education economics, namely those on the effects of exam schedules, intermediate exams, and exam resits on student performance. Without the ambition of being exhaustive, we report below some key findings from this literature.

Exam scheduling. A key reference for our study is Di Pietro (2013), which analyzes a reform in a UK university where the examination format was changed from end-of-semester exams to end-of-academic-year exams. Using a DiD strategy, the author compares mid-term exams, which remained unchanged during the study period, with final exams that were affected by the reform. The switch to end-of-academic-year final exams resulted in

hours) taught over one sub-cycle (5 weeks). Finally, the examination period was extended by adding one week in the winter exam session and one week in the summer exam session to allow time for the second intermediate exam.

a 6.7% decrease in student scores, equivalent to approximately -0.3 standard deviations (SD). A study on a medical school (Schmidt et al., 2022) compared the effects of two different exam scheduling regimes, exam bundling (back-to-back exams) versus spaced-out exams—on student performance. The authors found that switching from spaced-out exams to back-to-back exams resulted in declining student performance and overall satisfaction. Another recent study conducted in Greece by Goulas and Megalokonomou (2020), leverages quasi-random variation in secondary school exam schedules across cohorts, grades, and subjects, revealing that exam scheduling primarily impacts performance in STEM exams. In particular, the authors identify a positive effect of exam order ("warm-up effect"), where students perform better on later exams, and a negative effect ("fatigue effect"), where performance declines the longer the time interval since the first exam. These effects are more pronounced for higher-performing students. However, the authors estimate that optimizing the exam schedule would yield a modest improvement in overall performance (around (+0.02 SD).

Presence and frequency of intermediate tests. A randomized experiment conducted at an Italian university by De Paola and Scoppa (2011) compared the performance of students in an introductory economics course. One group was allowed to take an intermediate exam and was informed about the results obtained, whereas the control group could only take the final exam. The authors found that students who took the intermediate exam showed significant improvements in their final exam pass rates and grades, with the effects being more pronounced among higher-ability students. Interestingly, the positive outcomes were primarily attributed to the "workload division or commitment" mechanism rather than the feedback itself. Other studies have found positive effects of increasing the frequency of intermediate (or mid-term) tests on student performance (Sulun et al., 2018).

Exam resits/retakes. Several papers have focused on the effect of resits of high-stakes exams, such as those required for admission to selective higher education programs or medical schools, on performance in the same tests or entrance into higher education. Goodman et al. (2020) leveraging students' higher probability of retaking the SAT test due to left-digit bias, report positive effects both on SAT score and the probability of a four-year college entry in the US. Effects are larger for low-income students and under-represented minorities. Frisancho et al. (2016) document similar effects in the Turkish college entry exam. Retaking entails some cumulative learning and, therefore, an increase in exam scores, especially among less advantaged students. The results may depend on the frequency of the re-tests, and the benefits of re-testing on learning may be lower when a high number of retests is allowed or exams are lower stakes, such as single course exams at university. In general, students taking a higher number of retests are on average of lower ability (or exerting lower effort), with potentially lower gains (Cates, 1982; Rubright et al., 2022). Bizopoulou et al. (2022) found that students who retake national end-of-high-school high-stakes exams improved their performance by half a standard deviation but did not receive offers from higher-quality post-secondary placements. A theoretical foundation for these results is provided by Krishna et al. (2018). Although each student would be interested in retaking the exam to improve her score, this would not necessarily increase her probability of entering the higher education system because general equilibrium effects generate an increase in acceptance cutoffs. Another theoretical paper predicts that thanks to retakes, rational students improve their probability of passing exams, but with negative effects on student effort (Kooreman, 2013). This paper lays the foundation for empirical tests, demonstrating that the model's prediction of lower student effort is confirmed by simulated assessment data from laboratory experiments (Nijenkamp et al., 2016), and it is countervailed when study time investments quickly depreciate, such as when students forget what they previously studied (Nijenkamp et al., 2018). By contrast, the evidence was less consistent with the model when tested in a field experiment using actual student assessment data (Nijenkamp et al., 2022). Recent theoretical work by Bertola (2024), which incorporates in the model multiple retakes extending Kooreman (2013), confirms that introducing additional retakes generally increases the probability that students will pass an exam, but reduces the preparation for exams. According to this model, it is optimal for students to make an initial attempt without much preparation if they have the opportunity to better prepare for a retake after a failed exam.

What do previous empirical and theoretical studies suggest about the expected effects of the exam reform we are examining? The expected effects on student outcomes can go in either direction. On the one hand, by removing the exam sessions that were further in time from the end of the teaching periods, the reform should have produced a stronger incentive to sit exams immediately after the end of the teaching terms and to avoid the negative effects of exam procrastination (Kim and Seo, 2015). This may lead to lower depreciation of the knowledge acquired in the lectures, especially for attending students, with effects similar to those found in Di Pietro (2013).[15] As previously described, the exam reform mandated teachers to administer an intermediate exam at the end of each teaching sub-cycle. This measure could further improve student performance through "workload division or commitment" (De Paola and Scoppa, 2011). However, most teachers in the Faculty of Economics were already voluntarily conducting intermediate exams to encourage student attendance before the reform. Therefore, the additional positive effect on student performance from simply making these exams mandatory remains unclear.

On the other hand, the availability of multiple resits is key in the context we are studying. Based on the available literature, we should expect a decline in student progression after the reduction in the number of resits takes effect (Bertola, 2024; Kooreman, 2013). Bertola (2024) demonstrates that increasing the number of retakes, ceteris paribus, eventually raises a student's probability of passing exams but reduces student preparation (or level of competence), especially at earlier attempts. Moreover, reducing exam retakes also increases students' need for multi-tasking, i.e. simultaneously studying different subjects, with potential negative effects on performance (Lavy, 2023).

In this paper, because of the features of our research design, we mainly focus on student performance in the first year of studies, examining the effects of retakes on first-year exams within that time frame. Focusing only on first-year students enables us to abstract from the complications determined by the fact that considering the following academic years, laggard students may strongly differ in the combination of the exams they can take. Indeed, laggard students may retake exams on courses programmed in the academic year $t-1$ (or previous years), which they failed or did not attempt yet, also in the following years, t, $t+1$, and so on until they eventually graduate. There is no time limit for completing their educational path as long as they continue paying tuition fees. By concentrating on the first year of study, our main analysis captures the outcomes of early retakes especially. We provide nonetheless some exploratory evidence on long-term academic outcomes such as the probability of on-time graduation in Section 6.2.

In the Italian higher education system, retakes serve not only as remedial exams, providing additional opportunities for weaker students to pass them, but also as a strategy for students to improve or maintain their GPAs by refusing previously awarded grades. This GPA management is crucial, as higher final grades can enhance employment prospects after graduation (i.e., many jobs in public administration require a minimum GPA to qual-

[15] Although going in the opposite direction as in Di Pietro (2013) the reform determined a longer interval between subject teaching and assessment.

ify for competitive examinations). We expect the first aspect to be especially relevant to low-ability students and the second to medium-ability students, assuming that high-ability students are likely to pass exams and receive high grades in fewer (possibly their first) attempts. Consequently, fewer opportunities for refusing grades, coupled with larger spacing between the main session and resits, should increase student effort in their first exam attempts and potentially accelerate academic progression.

Finally, the reduction in the number of retakes may have produced a change in the student intake, resulting in a decrease in the number of lower- and intermediate-ability students, those who likely had the most to lose from fewer retake opportunities, either in terms of passing exams or improving their GPA. These selection effects are further explored in Subsection 6.1.3.

4 Data and sample selection

We use administrative data on students who enrolled in academic programs at the University of Bologna from 2007-08 onward, made available by the statistical office of the University of Bologna. The data provide rich information on students' academic careers, including the specific degree course attended, the registration date, the number of credits (ECTS) earned, exams taken, and average grades received in each calendar year. For students who graduated by 2019, it also includes the graduation date and final grade. Additionally, the data contain students' demographic information, such as date of birth, gender, nationality, region of residence at the time of enrolment, and information on their pre-collegiate preparedness, namely the upper secondary school track attended.

We restrict our analysis to students who enrolled in a bachelor's degree program (or course) between the 2007-08 and the 2013-14 academic year. Later cohorts are excluded because, as mentioned in Section 2, in the academic year 2014-15 an optional (on-demand) resit opportunity was introduced, partly counteracting the drastic change in exams' organization introduced in 2010. We identify students fully affected by the reform as those who enrolled in a degree course taught in the Faculty of Economics of the Bologna campus from the 2010-11 academic year onward. These correspond to eight degree programs, within the broad fields of study: "Social and behavioral sciences" and "Business, administration, and law," as categorized by the *Fields of Education and Training (FOET) 2013 classification*. The former includes courses in law and economics and economics and finance, while the latter includes business administration-related courses. To select a credible control group, we focus our analysis on degree programs within the same two FOET fields mentioned above. These correspond to ten degree programs, mainly related to political sciences and psychology.[16] We restrict the sample to students who were no older than 25 years at the time of enrolment. The final sample consists of $9,083$ students.

Our main outcomes of interest are measures of students' regularity in their academic careers at the end of their first year of enrolment. We define dropout as the probability of leaving UniBO's administrative registries after their first year of enrolment. We cannot track students who leave UniBO to enroll in similar or different degree programs at other universities.[17] Thus, our measure is an upper bound of the overall probability of university dropout. We assess students' academic progression by calculating the total number of ECTS accumulated and exams passed within the first year of the academic program. This information is recorded in UniBO student registries by calendar year rather than academic year. Thus, for each enrolment cohort, starting in the academic year $t/t+1$, we consider the number of credits and exams accumulated until December $t+1$. We mainly focus on first-year performance because the required exams to pass for students enrolled in the same degree program are the same during this period. In the second and subsequent years of study, the exams passed and credits earned are significantly influenced by the number of exams that students have already passed. This introduces complications in the analysis, as the number of available exams becomes endogenous in each period. For instance, students who fall behind (who do not pass first-year exams like mathematics) can theoretically earn more credits during the second year of their studies since they can still take all the exams they did not pass in the first year. However, since time to degree completion was an important determinant of the new exam policy, we also attempt to estimate long-term effects on student performance with this caveat in mind.

Table 2 presents the mean first-year outcomes for students not affected by the reform before and after the reform takes place (columns (1) and (4), respectively) and for students affected by the reform during the same periods (columns (2) and (5), respectively). Columns (3) and (6) show the differences across the two groups for the period before and after the reform. The first-year dropout rate for cohorts enrolled from 2007-08 to 2009-10 (before the reform) was approximately 13.5% for the control group and 13% for the

[16] Degree programs offered by the Faculty of Political Science and the Faculty of Economics at the Forlì campus are excluded since the reorganization of exams for these courses was implemented earlier.

[17] This is common in studies using data from a single institution (see. e.g., Minaya et al., 2022; Carrieri et al., 2015).

Table 2: Means of first-year student outcomes by treatment status before and after the reform

Variables:	Mean control t(0) (1)	Mean treated t(0) (2)	Diff t(0) (3)	Mean control t(1) (4)	Mean treated t(1) (5)	Diff t(1) (6)	DiD (7)	s.e. DiD (8)
University dropout	0.135	0.130	-0.005	0.126	0.0669	-0.059	-0.054***	(0.013)
No. ECTS 1st year	36.22	33.29	-2.927	39.91	48.05	8.148	11.075***	(0.861)
No. exams 1st year	4.463	4.701	0.239	4.642	5.870	1.228	0.990***	(0.112)
Observations	1973	2127		2740	2243			

*** p<0.01, ** p<0.05, * p<0.1
Notes: The table reports means of the student outcomes in the period before the reform (t(0)) and the period after the reform (t(1)), and the post- vs. pre-reform differences in columns (3) and (6), respectively. DiD estimates in column (7) and standard errors in column (8) are obtained using a linear regression excluding control variables estimated with OLS. "University dropout" refers to be probability of dropping out from the University of Bologna.

treated group. For the post-reform cohorts (2010-11 to 2013-14), it remained rather stable for the control group (12.6%), and it dropped to 6.7% for the treated group, leading to an estimated (DiD) effect of -5.4 percentage points. It is important to note that the dropout rates mentioned refer to students who left the University of Bologna. However, some of them might have continued their studies at other universities. While this may not significantly affect the University of Bologna's statistics, it has broader implications for the country's higher education system. When students transfer to other universities or decide to enroll in a different degree program (by changing their major), they incur both financial and time costs.[18] The number of first-year ECTS credits and exams changed marginally for non-treated students during the estimation period, while it significantly increased for treated students, with estimated DiD effects of approximately 11 ECTS credits and 1 exam, respectively.

5 Empirical Strategy

Our empirical strategy exploits the fact that the new exam policy applied to students registering in degree programs (courses) at the Faculty of Economics of Bologna from the academic year 2010-11 onward, while students registered before remained under the old regime (except for overdue exams for subjects taught in previous years, see Section 2).

[18] In case students transfer to other universities, even in a similar degree program (e.g., in the same major), the credits earned and exams passed are only partially recognized.

To evaluate the impact of the reform on student outcomes, we rely on DiD. Since all treated units start to be treated at the same time, and treatment is an absorbing state, our baseline DiD model can be estimated using the Two-Way Fixed-Effect estimator (TWFE):

$$Y_{ict} = \alpha + \gamma(Treat_c * Post_t) + \delta' DegreeCourse_c + \lambda' Cohort_t + \epsilon_{ict} \tag{1}$$

where Y_{ict} is the outcome of student i from cohort t enrolled in degree program c; $Treat_c$ is the treatment dummy that takes value one if the student is enrolled in a degree course subject to the new policy, which reduced the number of exam retakes from six to three (i.e. those belonging to the Faculty of Economics after 2010-11), and zero if the student is enrolled in a comparable program to economics, such as political science and sociology, which did not undergo the reform; $Post_t$ is a time indicator that takes value one for the academic years from 2010-11 onward and zero otherwise; $Cohort_t$ and $DegreeCourse_c$ are vectors of student cohort and degree courses fixed effects, respectively, and ϵ_{ict} is the error term. The coefficient γ of the interaction term ($Treat_c * Post_t$) captures the causal effect of interest (DiD estimate). We initially considered three outcomes of interest: i) an indicator for first-year dropout; ii) the number of exams taken in the first year; and iii) the number of credits (ECTS) earned in the same year.

We further enrich the model by estimating alternative specifications that include student-level controls:

$$Y_{ict} = \alpha + \gamma(Treat_c * Post_t) + \phi' \mathbf{X}_{ict} + \delta' DegreeCourse_c + \lambda' Cohort_t + \epsilon_{ict} \tag{2}$$

where $\mathbf{X}_{ict}$ is a vector of student-level characteristics, including an indicator for gender, nationality (Italian vs. foreign-born), region of residence upon university enrolment (Emilia Romagna — the administrative region in which Bologna is located — vs. other regions in Northern, Central or Southern and Insular Italy) and upper secondary school curriculum (academic humanities, science tracks, and vocational tracks).

The key assumption for any DiD strategy is that the outcomes in the treatment and control groups would follow the same time trend in the absence of treatment. Thus, the main threat to identification is a potential violation of the parallel trend assumption (PTA). One can consider two cases: i) the parallel trends assumption holds unconditionally (UPTA); or ii) it holds only after controlling for observed characteristics (CPTA). The

CPTA may potentially pose problems. Some papers show that the two-way fixed effect (TWFE) model, which is commonly used in the DiD analysis, does not recover the Average Treatment Effects on the Treated when time-varying covariates are included (Zeldow and Hatfield, 2021; Sant'Anna and Zhao, 2020). The idea is that the observed characteristics $\mathbf{X}$ themselves can be affected by the treatment (or their mean value may depend on the treatment), leading the TWFE to over-control (i.e. by including "bad" controls).

For this reason, we use as our baseline specification the unconditional (or unadjusted) DiD model as in equation (1), which assumes that the UPTA holds. A valid UPTA would be reassuring, as it allows us to estimate the overall effect of the reform (Zeldow and Hatfield, 2021). This includes both the "direct effect" on student performance resulting from behavioral changes or other effects (e.g. less knowledge depreciation) and the "indirect effect" on student performance mediated by compositional changes in student characteristics ($\mathbf{X}_{ict}$).

We test the UPTA by estimating an event-study DiD model as follows:

$$Y_{ict} = \alpha + \sum_{k=-2}^{-1} \beta_k \times Treat_c + \sum_{k=0}^{3} \beta_k \times Treat_c + \gamma' DegreeCourse_c + \lambda' Cohort_t + \epsilon_{ict}, \quad (3)$$

where Y_{ict}, $DegreeCourse_c$, and $Cohort_t$ are defined as above; β_k is a vector of coefficients capturing the lagged values in each academic year before 2010-11 (first summation) and leads values after the year 2010-11 (second summation); -2 and 3 are the lowest and highest number of lags and leads from the first introduction of the treatment (time 0) that we consider. Based on our data we have 3 periods before and 4 periods after treatment (time 0). Lag coefficients should be equal to zero, to guarantee the validity of the common trend assumption. In our analysis, the reference year is $2009-10$, the academic year immediately preceding the introduction of the new exam policy. The coefficients on the lead terms inform us about the potential time-varying impact of the exam policy change (our data allow us to examine effects up to the 2013-14 enrollment year, which corresponds to four periods after the treatment). Additionally, we report estimates of equation (3) including the vector $\mathbf{X}_{st}$ to check the validity of the conditional parallel trend assumption (CPTA).

Understanding the overall effect of the reform is crucial, but it is also important to determine whether potential changes in student performance are primarily driven by a selection effect, i.e., enrollment of higher-performing students, or by "virtuous" behaviors, i.e., the direct effects of the reform itself. We try to disentangle the two effects in various ways. First, in the spirit of mediation analysis, in addition to the unadjusted DiD, we also report the adjusted DiD estimates, in which we control for the potentially affected ("bad") control variables. This allows us to isolate the effect of the reform (β) from the effect mediated by changes in $\mathbf{X}_{ict}$. As a further strategy that addresses potential imbalances in the covariates before and after the reform (i.e., compositional changes), we adopt a matching-DiD analysis (Abadie, 2005). Finally, changes in the student intake induced by the reform, what we refer to as the "cream-skimming" effect, are interesting per se. To shed light on this, we: (i) use DiD analysis to examine changes in $\mathbf{X}_{st}$, specifically compositional shifts in the student intake; and (ii) apply the procedure proposed by Carrell et al. (2018) to determine the extent to which changes in performance can be attributed to the "cream-skimming" of students.

The results of our empirical analysis are discussed in the next section.

6 Results

In this section, we first provide evidence on the overall short-term effect of the reform on first-year academic performance. We then carry out a DiD event-study analysis to check the main identifying assumptions and investigate dynamic effects. We further show that the impact of compositional effects ("cream-skimming") is minimal; we run some additional robustness checks and provide some insights into the possible mechanisms through which the effect materialized. Finally, we provide some evidence on the long-term effects of the reform on student academic performance by investigating graduation outcomes.

6.1 Short-term effects on student performance: first-year outcomes

6.1.1 DiD baseline estimates

Table 3 presents the baseline estimates from equation (1), i.e. the unadjusted-DiD estimates. On average, the reform decreased the probability of dropping out by the end of the first year by 4.2 pp (-32%). This change is significant at the 5% level. The reform also caused an increase in the number of ECTS credits earned and of exams passed during the first year of studies by 11 (33%) and 1 (22%), respectively, with both changes statistically

Table 3: First-year outcomes: Unconditional DiD estimates

	University dropout (1)	No. ECTS 1st year (2)	No. exams 1st year (3)
Treat*Post	-0.042**	11.205***	1.061***
	(0.018)	(1.983)	(0.265)
R-squared	0.032	0.135	0.233
Cohort FE	Yes	Yes	Yes
Degree course FE	Yes	Yes	Yes
Mean outcome	.13	33.29	4.7
% effect	-32	33	22

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: The table reports the $Treat * Post$ interactions in equation (1), in which student characteristics are omitted. Standard errors clustered at the degree course by cohort level.

significant at the 1% level.[19]

As mentioned in the previous section, these results are obtained from a DiD specification that omits student characteristics, capturing both the indirect effect of the reform on student performance that is mediated by any compositional change in the average student characteristics, and the direct effect of the reform on performance when student characteristics are kept constant (e.g., due to changes in student behaviors).

To shed light on the magnitude of the direct effect of the reform, we present both the results from the estimation of a DiD specification including student characteristics (i.e. adjusted for covariates), and from a matching DiD in which the post-reform students are matched with those from the pre-reform period based on observable characteristics. Results are reported in Table 4 and Table 5, respectively. The adjusted-DiD and the matching-DiD estimates are only slightly smaller in magnitude, which is consistent with the limited "cream-skimming" effect documented in Subsection 6.1.3. In the adjusted DiD, we observe a 3.7 pp reduction in dropout, an increase of 10.49 ECTS credits, and a 0.97 increase in the number of exams passed in the first year. With matching DiD we find a 4.3 pp reduction in dropout, an increase of 10.14 ECTS credits, and a 0.92 increase in the number of exams passed during the first year.

Taken together, the results presented in this section indicate that the reform of the

[19] These estimates differ from those in Table 2 because the former also includes cohort and degree course fixed effects, while the latter only includes $Post_t$ and $Treat_c$ indicators variables.

Table 4: First-year outcomes: Conditional DiD estimates

	University dropout (1)	No. ECTS 1st year (2)	No. exams 1st year (3)
Treat*Post	-0.037**	10.486***	0.974***
	(0.018)	(1.952)	(0.244)
R-squared	0.072	0.221	0.307
Cohort FE	Yes	Yes	Yes
Degree course FE	Yes	Yes	Yes
Student-level covariates	Yes	Yes	Yes
Mean outcome	.13	33.29	4.7
% effect	-28	31	20

*** p<0.01, ** p<0.05, * p<0.1

Notes: The table reports the $Treat * Post$ interactions in equation (2), in which student characteristics are included. Standard errors clustered at the degree-course by cohort level.

Table 5: First-year outcomes: Matching-DiD estimates

	University dropout (1)	No. ECTS 1st year (2)	No. exams 1st year (3)
Treat*Post	-0.043**	10.142***	0.923***
	(0.021)	(2.061)	(0.271)
R-squared	0.066	0.219	0.293
Student-level covariates	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes
Degree course FE	Yes	Yes	Yes
Mean outcome	.13	33.29	4.7
% effect	-32	30	19

*** p<0.01, ** p<0.05, * p<0.1

Notes: The table reports the $Treat * Post$ interactions from the estimation of a kernel-based propensity score Matching DiD. Standard errors clustered at the degree course by cohort level.

organization of the exams had a positive impact on the performance of first-year students.

6.1.2 Event-study DiD

The main identifying assumption of our DiD approach is that the outcomes would have evolved in the same way for treated and control cohorts after the reform, had the reform not occurred. To gain insights into the credibility of this assumption, we investigate the presence of pre-reform differences in outcome trends across treated and control cohorts by estimating an event-study DiD model as described in equation (3). Importantly, we focus on the unconditional parallel trend assumption, as it would allow us to interpret the estimates in Table 3 as the overall effects of the new exam policy, including both the indirect effect on student characteristics and the direct effect on student performance (Zeldow and Hatfield, 2021; Minaya et al., 2022).

The results are presented in Figure 1. We take the first pre-reform cohort, i.e. 2009-10, as the baseline ($k = -1$) and plot the coefficients of the lag and lead interactions. The resulting coefficients are displayed for the main outcomes of interest, i.e. drop-out probability and the number of ECTS and exams in panels (a), (b), and (c), respectively. The vertical bars display the 90, 95, and 99% confidence intervals. The point estimates of the coefficients and their standard errors are reported in Table A1 in the Appendix.

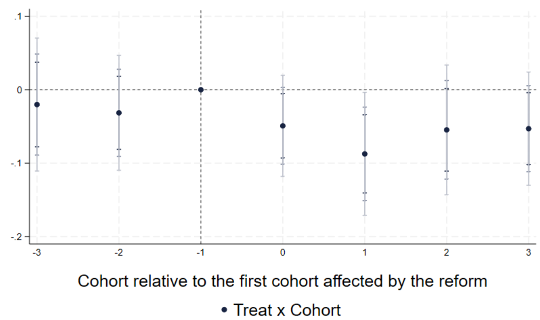
As it emerges from panel (a), the treated-control pre-reform cohorts' difference in first-year drop-out probability relative to the first pre-reform cohort is close to zero and is not statistically significant. The same difference is bigger in magnitude for post-reform cohorts, although it appears to be statistically significant only for the first and the last cohorts (at 1 and 10% significance level, respectively). The evidence for the other two outcomes of interest is more net: while the treated-control pre-reform cohorts' difference in the number of first-year ECTS earned (panel b) and exams passed (panel c) relative to the first pre-reform cohort is virtually zero, there is a clear jump for all post-reform cohorts. All in all, the results of the event-study DiD generally confirm the validity of the conditional parallel trend assumption.

Figure A1 in the Appendix displays the estimates of the event-study DiD with control variables.[20] The estimates of the treatment by cohort interactions are very close to those displayed in Figure 1.

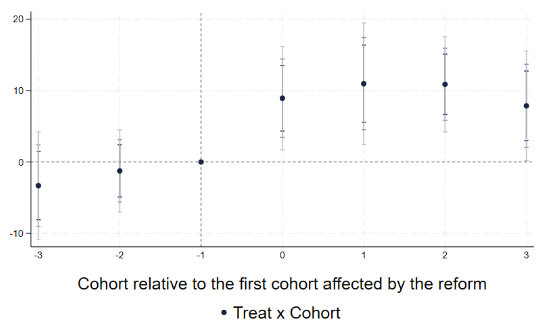
[20] Point estimates are reported in Table A2, in the Appendix.

Figure 1: First-year outcomes: Unconditional event-study DiD estimates

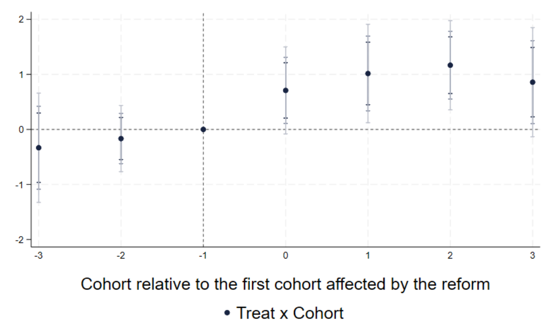
(a) University dropout



(b) No. ECTS 1st year



(c) No. exams 1st year



Notes: The figures display the estimated coefficients associated with the interaction between the $Treat$ indicator and cohort dummies relative to the first cohort affected by the exam reorganization (2010-11), omitting student characteristics (see equation (3)). The baseline is the cohort before the exam reorganization implementation. The dependent variables are the probability of university dropout by the end of the first year of studies (panel a); the number of ECTS accumulated during the first year of studies (panel b); the number of exams accumulated during the first year of studies (panel c). The regression contains student characteristics as displayed in Table A3 and degree courses fixed effects. Standard errors clustered at the degree course by cohort level.

6.1.3 "Cream-skimming" effects

A reduction in the number of retakes could lead to a "cream-skimming" effect by increasing the level of effort required of students to pass exams or reducing students' freedom in the allocation of their time.

To check for compositional changes in student intake, we do two things. First, we estimate a DiD model in which the dependent variables are now student characteristics, and then we compare students enrolled in reformed versus non-reformed degrees, both before and after the reform. We consider the same covariates, namely gender, nationality, region of residence upon enrolment, and upper secondary school track. This approach offers some insights but does not fully address whether the reform helps attract academically stronger students, increasing the average student performance. This is why we follow a second approach in line with Carrell et al. (2018). We regress student outcomes on students' characteristics using the *pre-reform* sample. The estimated coefficients represent the "weight" of each characteristic on the right-hand side in predicting student performance. We then apply these "weights" to the characteristics of *post-reform* students to predict their performance as if they were behaving in the pre-reform regime. In this new DID model, the observed outcome is replaced with the predicted one, dropping the control variables. The coefficient on $treated_i \times post_t$ is interpreted as the average increase in student performance attributed only to changes in student intake (the amount of "cream skimming" or student selection generated by the reform).

The more rational the students are, the more they consider their expected performance when choosing a degree program, and the larger we expect this effect to be.[21] Table A3 in the Appendix reports the results from the first analysis. While we see a drop in the percentage of enrolled women (-6 pp), the age at enrollment decreases marginally (-0.09 years). Changes in students' nationality and geographical origin are less pronounced, except for a drop in the number of students coming from the least performing regions, South and Islands (-3.5 pp). Quite interestingly, we observe a large increase in the percentage of students coming from the academic scientific track (*"liceo scientifico"*), by 10.3 pp, compensated by reductions in the percentages of students coming from the humanities academic and the vocational tracks, by 3.2 pp and 7 pp, respectively. The drop in the share of women in degrees affected by the reform seems inconsistent with a "cream-skimming"

[21] In fact, studies are questioning the ability of students to correctly predict their grades (Serra and DeMarree, 2016; Zafar, 2011) or their wages (Betts, 1996), and their rationality (DesJardins and Toutkoushian, 2005).

effect, as women perform better in higher education (Atzeni et al., 2022).[22] However, the increased selectivity of a degree program, driven by the reduction in the number of retakes, may have a discouraging effect on women, regardless of their ability (Saygin, 2016). The change observed in the student intake by secondary school track, on the other hand, is consistent with the idea that students recognize the higher effort needed to have a satisfactory performance in the degrees affected by the reform, leading to positive selection effects.

To better illustrate any selection effect, we report in Table A4 in the Appendix the results of a regression in which first-year outcomes of students in the *pre-reform* cohorts are regressed on gender, geographical origin, age at enrollment, secondary school track and cohort, and degree-course fixed effects. Females perform better on average (3.7 more credits and 0.5 more exams than men). Offsite students (*fuori sede*) generally outperform "local" students (i.e., those residing in Emilia Romagna at the time of enrollment), except for students from Southern and Insular regions, who earn 2.9 fewer credits and pass 0.4 fewer exams compared to local students. Foreign students perform worse than natives (with a penalty of 4 credits and 0.6 exams), although they are less likely to drop out (-4.5 pp). As expected students from the scientific track outperform the others.

Finally, Table A5 in the Appendix reports the estimates of our DiD (second analysis) based on Carrell et al. (2018). We see that selection explains only a small part of the observed improvements — approximately 0.8 additional credits earned and 0.1 more exams passed — with no notable effect on dropout rates.

6.1.4 Further specifications and robustness checks

Ministerial Decree on harmonising university course offerings. In 2004, the Italian government passed a reform (*Ministerial Decree 270/04*) aimed to reduce graduation times by standardizing course offerings across universities. Before the reform, all bachelor students were required to earn 120 credits to graduate, but universities had the flexibility to decide the number of exams students had to pass to achieve those credits. The reform thus limited all universities to a maximum of 12 courses for earning the necessary 120 credits. This was done by streamlining the curriculum so that exams covering similar topics were consolidated.[23] The law only required the universities to complete the reform

[22] See Table A4 in the Appendix, discussed below.

[23] See Malacrino et al. (2024) for evidence on the effects of this reform.

process by the academic year 2011. At the University of Bologna, the reform was implemented with reformed curricula for degree courses starting in the academic year 2008-09, for both courses under the Faculty of Economics and the ones included in the control group in our analysis. Thus, this change should not confound the effect of the main reform we are investigating in this paper. As a robustness check, we estimate our specification in equation (1) for the three short-term outcomes excluding the academic year in our sample prior to the Ministerial Decree implementation, i.e. 2007-08. The results are presented in column (1) of panels (a), (b), and (c) in Table A6 in the Appendix, corresponding to the three outcomes of interest. The estimated effects are very close to those reported in Table 3.

Other changes introduced by the reform. As illustrated in Section 2, the reorganization of exam sessions and schedules implemented at UniBO starting from 2010-11 was part of a broader set of actions, some of which were already implemented in the previous academic year (entry tests, reorganization of lecture schedules making them more compact, the extension of exam sessions by one week). Although the event-study coefficients reported in Figure 1 show that the largest difference in performance started to be evident since 2010-11, to rule out that the observed effects are driven by the above-mentioned changes rather than by the exam reorganization, we estimated equation (1) excluding the 2009-10 student cohort (donut-hole regression). The results from this empirical exercise – presented in columns (2) of panels (a), (b), and (c) in Table A6 in the Appendix – show that the DiD estimates are of similar magnitude as the baseline results. We also estimated a DiD for the pre-reform cohorts, imputing a fictitious treatment period where *Post* is an indicator for being enrolled for the first time in 2009-10, i.e. the year just before the actual introduction of the exam reform. Columns (3) of panels (a), (b), and (c) in Table A6 report the results of this falsification check. None of the coefficients of the interactions of the treatment indicator with the fictitious treatment period are significant. All in all, the changes that were already introduced in 2009-10 do not seem to have produced any "anticipation" effect by improving student performance, compared to the more substantial reorganization of exams put in place the following academic year.

Other changes implemented by UniBO. Another change in degree course curricula occurred in the year 2010-11, when the Faculty of Economics implemented a reorganization of the teaching offer. Before this change, six-degree courses in Economics were offered (Economics and Business, Economics and Profession, Economics and Law, Economics and

Finance, Economics and Marketing, Economics and Management). Starting from 2010-11, some of them were interrupted and merged with others,[24] and a brand-new program taught in English (Economics and Business) was introduced. Changes in degree supply across years are accounted for by degree course fixed effects in our main specification. As a further robustness check, we estimate equation (1) excluding the newly introduced degree in English as well as on a su-sample of treated degree courses that exist in both the pre and post-reform period (i.e., a balanced panel in terms of degree courses). The results are reported in columns (4) and (5), respectively, of all panels of Table A6 and show that our main findings are confirmed.

Control variables and clustering. Table A7 in the Appendix reports some additional robustness checks. Columns (1) and (3) report our baseline estimates without and with control variables, respectively. Columns (2) and (4) report estimates with clustering at the degree course level (instead of cohort*degree course level) and wild-bootstrap *p*-values to account for the low number of clusters without and with control variables, respectively. Finally, column (5) reports the doubly-robust DiD estimator proposed by Sant'Anna and Zhao (2020) (DRDiD, hereafter) which is particularly advised for models controlling for covariates.[25] The estimates on drop-out, in panel (a), show that clustering at the degree level, irrespective of applying wild-bootstrap or not, the effect ceases to be statistically significant at conventional levels. By contrast, the effect is larger and statistically significant when applying the doubly-robust estimator and amounts to a 5.7 pp reduction in the probability of dropout. As for the number of credits earned and the number of exams passed, in panel (b) and (c), respectively, clustering does not influence the precision of the estimates. The DRDiD estimates are very close to those obtained with TWFE.

6.1.5 Mechanisms

Teacher leniency or changing incentives to refuse low grades

A better student performance could be determined by higher teaching leniency (i.e., grade inflation) after the reform. Although this option was available also before the reform

[24] Economics and Profession was discontinued and integrated into Economics and Business, Economics and Law was discontinued and integrated into Economics and Finance, Economics and Management and Economics and Marketing were discontinued and integrated into the newly constituted degree course Management and Marketing

[25] We estimate the improved doubly robust DiD estimator based on the inverse probability of tilting and weighted least squares and report wild-bootstrap standard errors.

and probably had higher benefits for professors before the reform given the higher number of possible attempts,[26] one may put forward that professors realized after the reform that it was more difficult for students to pass exams, and as a consequence, they relaxed grading standards. Unfortunately, our data do not gather information on single exams. However, some information limited to the 2009-10, 2010-11, and 2011-12 cohorts is available thanks to the Fondazione Giovanni Agnelli (see Aina et al., 2021). Figure 2 shows the exam grade distribution for the three exam sessions. In principle, if a higher leniency were applied after the reform, we would expect a higher bunching around the grade of 18 — the minimum passing grade — especially in the September session (i.e., the last of the academic year). However, this is not observed in the data; bunching at 18 is generally lower after the reform, and this is particularly evident in the September session of 2011. The data show a slight increase in bunching at 18 only in the Winter session. It is worth noting, however, that bunching at 18 might be produced both by higher leniency and by higher students' incentives not to refuse low exam grades after the reform since it entailed higher waiting costs owing to the longer spacing between retakes. In the next section, we will try to disentangle the two effects by focusing on long-term academic outcomes, namely the final graduation mark, which partly depends on the exam GPA and partly on the mark obtained in the final dissertation. The idea is that if students were less likely to refuse low grades in exams after the reform, we should also observe lower final graduation marks.

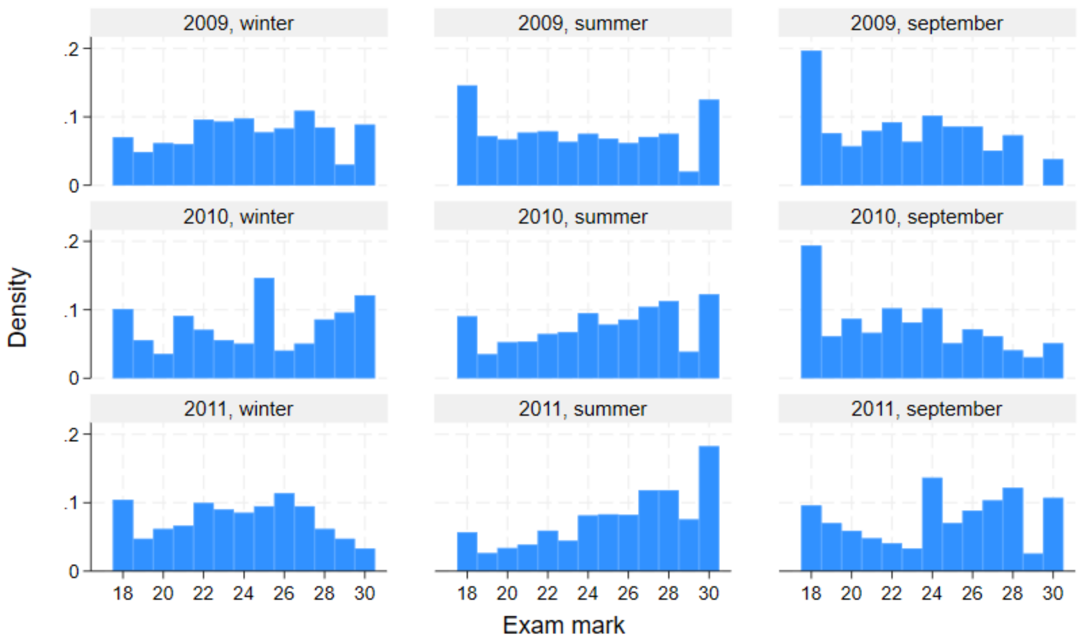
In Table 6, to further investigate the leniency hypothesis (or the lower incentive to refuse grades that is observational equivalent), we present a regression analysis on student number of exams and number of credits per semester of the lectures/exam session, for the sub-sample of students for whom exams data from Fondazione Giovanni Agnelli are available. First, columns (1) to (3) of panel (a) show that our main findings are confirmed in this sub-sample for first-year dropout, while are of smaller magnitude and less precisely estimated for first-(calendar) year number of exams and credits.[27] Columns (4) and (5) further show that results remain unchanged when focusing on exams passed only in the three exam sessions (until September) of the first academic year.

Panel (b) reports the results on the number of exams and credits separately by exam session and teaching semester of subjects. It appears that students tend to accumulate more exams and credits in the "regular" sessions right after the end of lectures and less

[26] In Italy, exam marking is usually done by the same professor teaching the course.

[27] In the main data used for this analysis, the Ministry of University and Research (MUR) provided all student performance indicators by calendar year and not by academic year.

Figure 2: Descriptive evidence on grades distribution



Notes. Source: ad hoc extraction from administrative data collected and released by the Italian Ministry of University and Research.

Table 6: Analysis by exam session: Unconditional DiD estimates

Panel (a). First-year outcomes on sub-sample (MUR data)

	University dropout (1)	No. exams at 31/12 (2)	No. ECTS at 31/12 (3)	No. exams at 30/09 (4)	No. ECTS at 30/09 (5)
Treat*Post	-0.050**	0.334	5.807*	0.423**	6.454**
	(0.020)	(0.197)	(3.061)	(0.190)	(3.027)
Observations	3,575	3,575	3,575	3,575	3,575
R-squared	0.031	0.266	0.126	0.254	0.101
Cohort FE	Yes	Yes	Yes	Yes	Yes
Degree course FE	Yes	Yes	Yes	Yes	Yes
Mean outcome	.14	4.69	35.31	4.24	31.96
% effect	-35	7	16	9	20

Panel (b). Results by exam session

	First semester courses				Second semester courses			
	Winter session		September session		Summer session		September session	
	Exams (1)	ECTS (2)	Exams (3)	ECTS (4)	Exams (5)	ECTS (6)	Exams (7)	ECTS (8)
Treat*Post	0.177	2.199**	-0.218***	-1.956***	1.235***	12.940***	0.010	0.413
	(0.123)	(0.861)	(0.048)	(0.490)	(0.133)	(1.967)	(0.036)	(0.437)
Observations	3,575	3,575	3,575	3,575	3,575	3,575	3,575	3,575
R-squared	0.270	0.194	0.063	0.066	0.345	0.262	0.105	0.128
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Degree course FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean outcome	1.76	11.22	.24	1.68	1.33	11.96	.31	2.96
% effect	10	19	-90	-116	92	108	3	13

*** p<0.01, ** p<0.05, * p<0.1

Notes: The sample only includes the 2009-10, 2010-11 and 2011-12 cohorts from the Ministry of University and Research (MUR) data. The table reports the Treat*Post interactions for different specifications. In panel (a) we report the results on our main outcomes in this sub-sample. In panel (b) we report the results on the number of exams and ECTS separately by exam session and teaching semester of subjects. Standard errors clustered at the degree course by cohort level.

in the resit session in September. As for the grades distribution of Figure 2, this could be determined by better performance (e.g., higher grades) or a lower incentive to refuse low grades, which both entail earning more credits in early attempts. As we anticipated, some evidence to distinguish between the two hypotheses is provided in the next section.

6.2 Long-term effects on student performance: graduation outcomes

Given the significant boost in credits earned in the first year, students affected by the reform may be more prone to graduate sooner than those under the old system. This aligns with previous research indicating that strong early academic performance predicts later success (Delogu et al., 2024). Moreover, graduating on time can have significant implications for students' early working careers. Previous research has shown that delayed graduation can negatively affect employability and early earnings (Garibaldi et al., 2012).

For this reason, in the current section, we offer some exploratory evidence on long-term academic outcomes related to graduation. As we stressed in Section 2, if we consider student performance beyond the first year, some of the student cohorts enrolled before 2010-11 — the first year of the reform— are partly treated for the exams of subjects of academic years preceding 2010-11.[28] This typically happens for laggard students, i.e., those who did not pass all the exams of a given academic year. For this reason, if the reform was effective, we should observe a violation of the parallel trend assumption, with student outcomes improving for the pre-reform cohorts closer to the 2010-11 cohort. With this caveat in mind, we now analyze graduation outcomes.

We define two dependent variables: graduation, an indicator that equals one if an individual graduates within our observation period and zero otherwise; and on-time graduation, which equals one if a student graduates within the standard duration of the degree program and zero otherwise.

Results are reported in columns (1) and (2) of Table 7. The DiD estimate for the treatment effect (Treat*Post) suggests that the reform has positively impacted the likelihood of graduation by approximately 5.7 pp, a 7% at baseline (column (1)). In the event study plotted in the top panel of Figure 3, although post-reform interactions are positive (except for the last one), they are imprecisely estimated. As expected, there is evidence that later pre-reform cohorts perform better, although not significantly, than the 2007-08 cohort, which, as far as graduation outcomes are concerned, is the only "pure" control.

Considering on-time graduation (column (2)), we see that the effects are more pronounced. The DiD estimate indicates that the reform led to a 9.8 pp increase in the probability of students graduating on time (significant at 1% level), which amounts to a 22% increase, demonstrating a substantial positive impact of the policy. Additionally, the

[28] For instance, students enrolled in 2009-10 had only three attempts for exams of the first-year sat in 2010-11 and the following academic years.

event-study analysis reveals positive effects that show up for cohorts enrolled starting from the academic year 2011-12 (central panel of Figure 3 and column (2) of Appendix Table A9), suggesting that it may take some time for students to reorganize their study plan or for professors to adjust to the new regime.[29]

All in all, this analysis shows that the reform increased both the probability of graduation and on-time graduation, significantly reducing the time the average student spent in education to get an undergraduate degree.

A major concern among students was that by making it more costly to refuse grades, the reform might have improved on-time graduation rates at the expense of lowering their final grades. This is relevant in the Italian context, as the final grade often determines access to high-quality jobs. In columns (2) to (6) of Table 7, we report the effect of the reform on the final grade and the probability of graduating with distinction (*cum laude*) both for all students who graduated and only for those who graduated on time. We observe an increase in the average final grade of 3.3 points (3%) and 2.8 points (2%),[30] respectively, for the two sub-samples, and no significant effect on the probability of getting a distinction. On the one hand, this last piece of evidence is consistent with the expectation that the reform was unlikely to be "binding" for top performers, who likely passed their exams on the first attempt and achieved good grades even before the reform. On the other hand, our analysis shows that the reduction in graduation times did not come at the cost of lower final marks for both top performers (those who graduate on time) and on average.

The final graduation mark is set according to a formula weighting both exam GPA and the score attributed to the final dissertation. Table A11 in the Appendix focuses on the effects only on the exam GPA. The effects are still positive but much lower than those on the final grade. This analysis allows us to exclude that professors might have compensated for the potential negative effects on the GPA, which instead increased after the reform, with a more lenient evaluation of the dissertations. Yet, on average, after the reform, treated students were getting better dissertation evaluations.

6.2.1 The role of family SES

Finally, in Tables A12 and A13 in the Appendix, we investigate potential winners and losers from the reform using unconditional and conditional DiD models, respectively. Unfortu-

[29] Conditional event-study DiD estimates are instead reported in Table A10.

[30] The final graduation mark varies between 66 and 110 *cum laude*.

Table 7: Graduation outcomes: Unconditional DiD estimates

			If graduated		If graduated on time	
	Graduation	On-time graduation	Final grad. mark	*Cum laude*	Final grad. mark	*Cum laude*
	(1)	(2)	(3)	(4)	(5)	(6)
Treat*Post	0.057**	0.091***	3.261***	0.009	2.789***	0.010
	(0.025)	(0.034)	(0.911)	(0.021)	(0.737)	(0.028)
Observations	9,083	9,083	6,641	6,641	4,660	4,660
R-squared	0.044	0.084	0.117	0.036	0.078	0.035
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Degree course FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean outcome	.73	.41	92.67	.07	98.39	.12
% effect	7	22	3	12	2	8

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

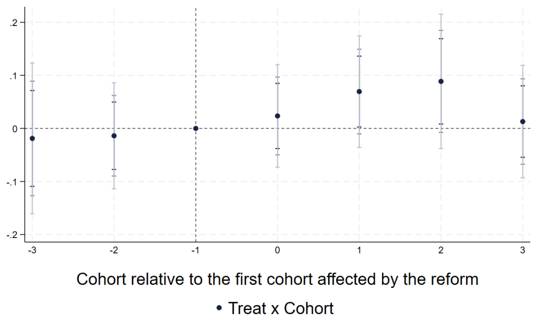
Notes: The table reports estimates for the Treat*Post when using longer-term outcomes: graduation, on-time graduation, final grade and cum laude if graduated and final grade and cum laude if graduated on time. Standard errors clustered at the degree course by cohort level.

nately, the available data do not provide information on students' family backgrounds for cohorts enrolled before 2015, making meaningful comparisons difficult. Yet, in Italy, there is a well-known social stratification in the choice of the upper secondary school track. High socio-economic status (SES) individuals tend to enroll in the academic tracks and those from low SES in vocational tracks (e.g., Triventi et al., 2021).[31] Thus, we can gain insights into the varying effects of the reform by SES by considering interactions between the treatment indicator and the academic track in upper secondary education. Reassuringly, the reform does not seem to have primarily benefited high-SES students. On the contrary, the latter seems to have gained less from the reduction in the number of retakes compared to those who came from vocational schools, although the interactions are never statistically significant at conventional levels. This is important as the greater flexibility allowed to students by a high number of retakes before the reform might have been especially useful to low-SES students, who were the ones more likely to work intensively while studying (Triventi, 2014).

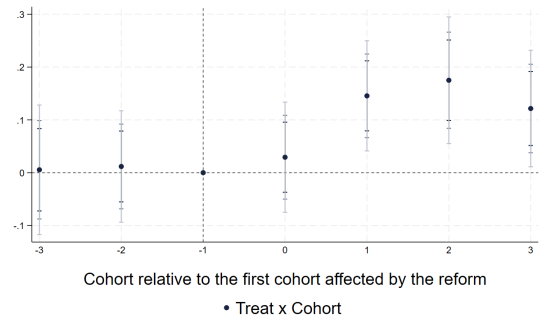
[31] With the data that we have, we checked the correlation between mother's or father's education and child high school track (academic vs. vocational) and the results are in line with the literature.

Figure 3: Graduation outcomes: Unconditional event-study DiD estimates

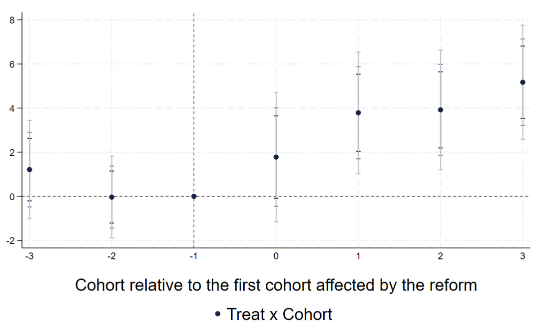
(a) Graduation



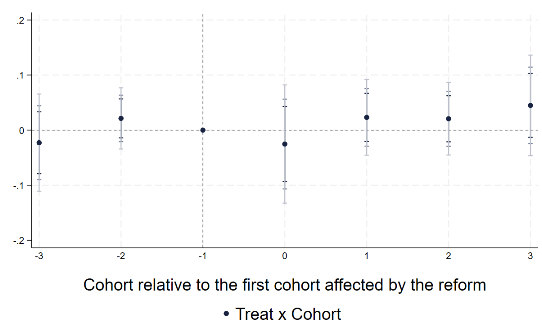
(b) On-time graduation



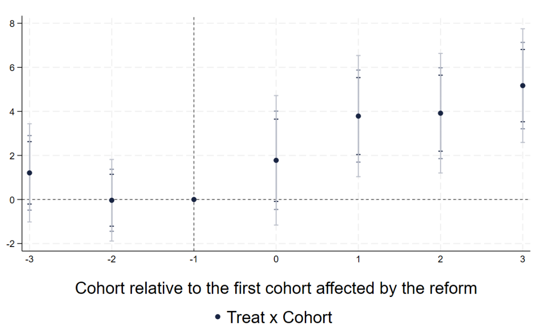
(c) Final grade if graduated



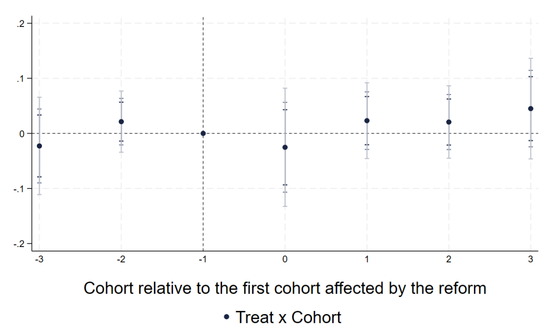
(d) *Cum laude* if graduated



(e) Final grade if graduated on time



(f) *Cum laude* if graduated on time



Notes: The figures display the estimated coefficients associated with the interaction between the *Treat* indicator and cohort dummies relative to the first cohort affected by the exam reorganization (2010-11), controlling for student characteristics. The baseline is the cohort before the exam reorganization implementation. The dependent variables are an indicator for being female (panel a); an indicator for being born abroad (panel b); an indicator for having moved region to study at the University of Bologna (panel c); an indicator for having graduated from an academic high school track (panel d). The regression includes cohort and degree courses fixed effects. Standard errors clustered at the degree course by cohort level.

7 Conclusions

The organization of university exams may affect student effort or have a direct impact on their academic performances. In this paper, we assess whether the extreme flexibility allowed to Italian university students to organize their studies, such as the very high number of exam resits, may be responsible for some undesirable outcomes, such as student dropout and slow academic progression, in turn determining graduation delays, which are very common in Italy. To this end, we leverage a reform that was introduced in the academic year 2010-11 by the Faculty of Economics of the University of Bologna, which reduced the number of examination attempts available to students by 50% (i.e., from 6 to 3) creating an incentive for students to sit exams just after the end of course lectures.

Using a difference-in-differences design, our paper demonstrates that the reform led to better first-year student performance, reducing dropout by 4.2 pp (−32%, i.e. about one-third) and increasing the number of earned credits and exams passed by about 11 ECTS (33%) and one exam (22%), respectively. This was accompanied by an improvement in the average quality of the student intake, which, however, accounted only for a tiny portion of the overall positive effect of the reform. Results are robust to a wide battery of robustness checks to address potential confounding factors such as institutional changes that pre-dated the exam reform. Focusing on performance in single exam sessions, we do not find any evidence that suggests that rising grade inflation was a major determinant of increased student performance.

Shifting the analysis to a longer time horizon, we document an increase in the probability of graduation of 5.7 pp (7%) and of on-time graduation of 9.1 pp (22%). Moreover, we do not observe a reduction in the final GPA or student graduation marks. This is important as the fear that the lower number of attempts would have impaired student performance was one of the major arguments pushing students to oppose the reform. Importantly, a heterogeneity analysis by a proxy of family background, namely having attended a vocational track in secondary education, shows that low-SES, those who were more likely to work, benefited from the reform.

Our analysis suggests that reforming the organization of university assessments may represent a powerful policy instrument to boost students' educational outcomes and tackle the age-old problem of Italian university students' very long graduation times.

Although this paper features one of the few quasi-experimental assessments of a reduction in the number of exam retakes on student performance, it has nonetheless the limitations of studies focusing on single institutions. For several reasons, educators may not necessarily expect similar effects if the same reform is implemented in other higher education institutions or countries. First, Italy is characterized by a very high number of exam retakes. Thus, their reduction may not necessarily negatively impact student performance or even improve it, like in our case, imposing some "discipline" on the least self-organized students. Second, UniBO is a very prestigious institution attracting better-than-average students. For future research, it would be interesting to investigate whether similar reforms implemented in less selective institutions confirm our main findings.

References

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies 72*(1), 1–19.

Aina, C., M. Bratti, and E. Lippo (2021). Ranking high schools using university student performance in italy. *Economia Politica 38*(1), 293–321.

Atzeni, G., L. G. Deidda, M. Delogu, and D. Paolini (2022). Drop-out decisions in a cohort of Italian universities. In D. Checchi and T. Jappelli (Eds.), *Teaching, Research and Academic Careers*, pp. 71–103. Springer, Cham.

Bertola, G. (2024). Retake opportunities, pass probabilities and preparation for exams. *Research in Economics*.

Betts, J. R. (1996). What do students know about wages? Evidence from a survey of undergraduates. *Journal of Human Resources*, 27–56.

Bizopoulou, A., R. Megalokonomou, and S. Simion (2022). Do second chances pay off? Evidence from a natural experiment with low-achieving students. *CESifo Working Paper No. 9620*.

Booij, A. S., E. Leuven, and H. Oosterbeek (2017). Ability peer effects in university: Evidence from a randomized experiment. *The review of economic studies 84*(2), 547–578.

Carrell, S. E., M. Hoekstra, and E. Kuka (2018). The long-run effects of disruptive peers. *American Economic Review 108*(11), 3377–3415.

Carrieri, V., M. D'Amato, and R. Zotti (2015). On the causal effects of selective admission policies on students' performances: Evidence from a quasi-experiment in a large Italian university. *Oxford Economic Papers 67*(4), 1034–1056.

Cates, W. M. (1982). The efficacy of retesting in relation to improved test performance of college undergraduates. *The Journal of Educational Research 75*(4), 230–236.

De Paola, M. and V. Scoppa (2011). Frequency of examinations and student achievement in a randomized experiment. *Economics of Education Review 30*(6), 1416–1429.

Delogu, M., R. Lagravinese, D. Paolini, and G. Resce (2024). Predicting dropout from higher education: Evidence from italy. *Economic Modelling 130*, 106583.

DesJardins, S. L. and R. K. Toutkoushian (2005). Are students really rational? The development of rational thought and its application to student choice. In *Higher education: Handbook of theory and research*, pp. 191–240. Springer.

Di Pietro, G. (2013). Exam scheduling and student performance. *Bulletin of Economic Research 65*(1), 65–81.

Feld, J. and U. Zölitz (2017). Understanding peer effects: On the nature, estimation, and channels of peer effects. *Journal of Labor Economics 35*(2), 387–428.

Frisancho, V., K. Krishna, S. Lychagin, and C. Yavas (2016). Better luck next time: Learning through retaking. *Journal of Economic Behavior & Organization 125*, 120–135.

Garibaldi, P., F. Giavazzi, A. Ichino, and E. Rettore (2012). College cost and time to complete a degree: Evidence from tuition discontinuities. *Review of Economics and Statistics 94*(3), 699–711.

Goodman, J., O. Gurantz, and J. Smith (2020). Take two! SAT retaking and college enrollment gaps. *American Economic Journal: Economic Policy 12*(2), 115–58.

Goulas, S. and R. Megalokonomou (2020). Marathon, hurdling, or sprint? the effects of exam scheduling on academic performance. *The BE Journal of Economic Analysis & Policy 20*(2).

Jones, T. R., D. Kreisman, R. Rubenstein, C. Searcy, and R. Bhatt (2022). The effects of financial aid loss on persistence and graduation: A multi-dimensional regression discontinuity approach. *Education Finance and Policy 17*(2), 206–231.

Ketel, N., J. Linde, H. Oosterbeek, and B. van der Klaauw (2016). Tuition fees and sunk-cost effects. *The Economic Journal 126*(598), 2342–2362.

Kim, K. R. and E. H. Seo (2015). The relationship between procrastination and academic performance: A meta-analysis. *Personality and Individual Differences 82*, 26–33.

Kooreman, P. (2013). Rational students and resit exams. *Economics Letters 118*(1), 213–215.

Krishna, K., S. Lychagin, and V. Frisancho (2018). Retaking in high stakes exams: Is less more? *International Economic Review 59*(2), 449–477.

Lavy, V. (2023). The effect of multitasking on educational outcomes and academic dishonesty. Technical report, National Bureau of Economic Research.

Malacrino, D., S. Nocito, and R. Saggio (2024). Do reforms aimed at reducing time to graduation work? evidence from the italian higher education system. Technical report, National Bureau of Economic Research.

Michaelis, J. and B. Schwanebeck (2016). Examination rules and student effort. *Economics Letters 145*, 65–68.

Minaya, V., T. Agasisti, and M. Bratti (2022). When need meets merit: The effect of increasing merit requirements in need-based student aid. *European Economic Review*, 104164.

Nijenkamp, R., M. R. Nieuwenstein, R. de Jong, and M. M. Lorist (2016). Do resit exams promote lower investments of study time? Theory and data from a laboratory study. *PLoS One 11*(10), e0161708.

Nijenkamp, R., M. R. Nieuwenstein, R. de Jong, and M. M. Lorist (2018). Controlling the resit effect by means of investment depreciation. *Journal of Cognition 1*(1).

Nijenkamp, R., M. R. Nieuwenstein, R. De Jong, and M. M. Lorist (2022). Second chances in learning: Does a resit prospect lower study-time investments on a first test? *Journal of Cognition 5*(1).

OECD (2008). *Education at a glance 2008: OECD indicators*. Organisation for Economic Co-operation and Development (OECD), Paris.

Rubright, J. D., T. Q. Ong, M. G. Jodoin, D. A. Johnson, and M. A. Barone (2022). Revisiting retake policy: Analyzing the success rates of examinees with multiple attempts on the United States medical licensing examination. *Academic Medicine 97*(8), 1219–1225.

Sant'Anna, P. H. and J. Zhao (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics 219*(1), 101–122.

Saygin, P. O. (2016). Gender differences in preferences for taking risk in college applications. *Economics of Education Review 52*, 120–133.

Schmidt, M., B. Pinney, and S. Werning (2022). Declining student performance and satisfaction following back-to-back scheduling of foundational science exams: Experience at a large us osteopathic medical school. *Education Sciences 12*(2), 94.

Scott-Clayton, J. (2011). On money and motivation: A quasi-experimental analysis of financial incentives for college achievement. *Journal of Human resources 46*(3), 614–646.

Serra, M. J. and K. G. DeMarree (2016). Unskilled and unaware in the classroom: College students' desired grades predict their biased grade predictions. *Memory & Cognition 44*(7), 1127–1137.

Stinebrickner, R. and T. R. Stinebrickner (2008). The causal effect of studying on academic performance. *The BE Journal of Economic Analysis & Policy 8*(1).

Sulun, E., E. Nalbantoglu, and E. K. Oztug (2018). The effect of exam frequency on academic success of undergraduate music students and comparison of students performance anxiety levels. *Quality & Quantity 52*(1), 737–752.

Triventi, M. (2014). Does working during higher education affect students' academic progression? *Economics of education review 41*, 1–13.

Triventi, M., C. Barone, and M. Facchini (2021). Upper secondary tracks and student competencies: A selection or a causal effect? evidence from the italian case. *Research in Social Stratification and Mobility 76*, 100626.

Zafar, B. (2011). How do college students form expectations? *Journal of Labor Economics 29*(2), 301–348.

Zeldow, B. and L. A. Hatfield (2021). Confounding and regression adjustment in difference-in-differences studies. *Health Services Research 56*(5), 932–941.

A Appendix

Tables

Table A1: First-year outcomes: Unconditional event-study DiD estimates

	University dropout	No. ECTS 1st year	No. exams 1st year
	(1)	(2)	(3)
Treat==1 & Yr==2008	-0.031	-1.248	-0.166
	(0.030)	(2.226)	(0.233)
Treat==1 & Yr==2010	-0.049*	8.924***	0.708**
	(0.027)	(2.802)	(0.306)
Treat==1 & Yr==2011	-0.087***	10.955***	1.016***
	(0.032)	(3.295)	(0.347)
Treat==1 & Yr==2012	-0.055	10.872***	1.167***
	(0.034)	(2.579)	(0.314)
Treat==1 & Yr==2013	-0.053*	7.858**	0.858**
	(0.030)	(2.970)	(0.384)
Cohort FE	Yes	Yes	Yes
Degree course FE	Yes	Yes	Yes

*** p<0.01, ** p<0.05, * p<0.1

Notes: The table reports the $Treat * Year$ interactions of a DiD equation that excludes student characteristics. Standard errors clustered at the degree-course by cohort level.

Table A2: First-year outcomes: Conditional event-study DiD estimates

	University dropout	No. ECTS 1st year	No. exams 1st year
	(1)	(2)	(3)
Treat==1 & Yr==2008	-0.034	-0.474	-0.078
	(0.028)	(2.163)	(0.204)
Treat==1 & Yr==2010	-0.050**	9.321***	0.755***
	(0.024)	(2.615)	(0.267)
Treat==1 & Yr==2011	-0.080**	10.357***	0.943***
	(0.031)	(3.491)	(0.352)
Treat==1 & Yr==2012	-0.052*	10.665***	1.143***
	(0.031)	(2.488)	(0.285)
Treat==1 & Yr==2013	-0.045	7.318**	0.797**
	(0.028)	(2.870)	(0.363)
Student-level covariates	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes
Degree course FE	Yes	Yes	Yes

*** p<0.01, ** p<0.05, * p<0.1

Notes: The table reports the $Treat * Year$ interactions of a DiD equation that includes student characteristics. Standard errors clustered at the degree-course by cohort level.

Table A3: Student characteristics before vs. after the reform

Variables:	Mean control t(0)	Mean treated t(0)	Diff t(0)	Mean control t(1)	Mean treated t(1)	Diff t(1)	Diff-in-diff	s.e. DiD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	0.605	0.457	-0.148	0.607	0.399	-0.208	-0.060***	(0.021)
Age at enrolment	19.52	19.35	-0.171	19.56	19.29	-0.263	-0.091**	(0.043)
Foreign born	0.0634	0.117	0.0537	0.0715	0.111	0.0390	-0.015	(0.012)
Region of residence:								
Emilia-Romagna	0.606	0.720	0.114	0.548	0.674	0.126	0.012	(0.020)
Other North	0.123	0.0475	-0.0752	0.127	0.0624	-0.0646	0.011	(0.012)
Centre	0.109	0.0625	-0.0464	0.141	0.0999	-0.0410	0.005	(0.013)
South and Islands	0.143	0.130	-0.0127	0.168	0.120	-0.0475	-0.035**	(0.015)
Abroad	0.0193	0.0395	0.0202	0.0161	0.0432	0.0272	0.007	(0.007)
High school curriculum:								
academic humanities	0.374	0.142	-0.231	0.421	0.157	-0.263	-0.032*	(0.018)
academic science	0.313	0.428	0.115	0.323	0.541	0.218	0.103***	(0.020)
vocational & others	0.314	0.430	0.116	0.256	0.301	0.0455	-0.070***	(0.020)
Observations	1973	2127		2740	2243			

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: The table reports, for each covariate: the average value in the pre-reform period for the control and treated groups (columns 1 and 2, respectively) and their difference (column 3); the average value in the post-reform period for the control and treated groups (columns 4 and 5, respectively) and their difference (column 6); the difference in differences obtained and its standard error (columns 7 and 8, respectively) obtained from a simple DiD specification estimated for each covariate.

Table A4: First-year performance regressions (pre-reform student cohorts)

	University dropout	No. ECTS 1st year	No. exams 1st year
	(1)	(2)	(3)
Female	-0.008	3.735***	0.482***
	(0.012)	(0.669)	(0.091)
Foreign born	-0.045**	-4.053**	-0.582**
	(0.022)	(1.585)	(0.240)
Age at enrolment	0.059***	-3.909***	-0.497***
	(0.006)	(0.367)	(0.039)
Region of residence:			
Other North	-0.007	1.042	0.104
	(0.027)	(1.424)	(0.161)
Centre	-0.012	1.545	0.144
	(0.013)	(1.075)	(0.136)
South and Islands	-0.015	-2.922***	-0.351***
	(0.013)	(0.917)	(0.101)
Abroad	-0.044	-0.765	-0.084
	(0.034)	(2.229)	(0.288)
High school curriculum:			
academic science	-0.034**	2.235**	0.294**
	(0.013)	(0.891)	(0.112)
vocational & others	0.070***	-5.026***	-0.577***
	(0.021)	(1.217)	(0.139)
Observations	4,100	4,100	4,100
R-squared	0.085	0.179	0.271
Cohort FE	Yes	Yes	Yes
Degree course FE	Yes	Yes	Yes

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: The table reports the OLS estimates of student performance equations (with different student outcomes) in the pre-reform period on students' observed characteristics. Standard errors clustered at the degree course by cohort level.

Table A5: First-year outcomes: “Cream-skimming” effect

	University dropout	No. ECTS 1st year	No. exams 1st year
	(1)	(2)	(3)
Treat*Post	-0.007	0.769	0.098
	(0.006)	(0.490)	(0.061)
R-squared	0.300	0.432	0.637
Cohort FE	Yes	Yes	Yes
Degree course FE	Yes	Yes	Yes

*** p<0.01, ** p<0.05, * p<0.1

Notes: The table reports the $Treat * Post$ interactions of a DiD equation using the predicted student outcomes as dependent variables. The latter are predicted using a linear regression estimated on the pre-reform period (see Table A4). Standard errors clustered at the degree course by cohort level.

Table A6: First-year outcomes: Robustness checks

	No 2007	No 2009	Fake treat.	No English course	Balanced sample
	(1)	(2)	(3)	(4)	(5)
(a) University dropout 1st year					
Treat*Post	-0.050**	-0.031		-0.042**	-0.040**
	(0.019)	(0.021)		(0.018)	(0.019)
Treat*FakePost			0.023		
			(0.025)		
Observations	7,638	7,703	4,100	8,721	3,904
Cohort FE	Yes	Yes	Yes	Yes	Yes
Degree course FE	Yes	Yes	Yes	Yes	Yes
(b) No. ECTS 1st year					
Treat*Post	10.701***	12.379***		10.832***	10.923***
	(2.138)	(2.245)		(1.911)	(1.923)
Treat*FakePost			2.395		
			(2.081)		
Observations	7,638	7,703	4,100	8,721	3,904
Cohort FE	Yes	Yes	Yes	Yes	Yes
Degree course FE	Yes	Yes	Yes	Yes	Yes
(c) No. exams 1st year					
Treat*Post	1.059***	1.204***		1.021***	1.012***
	(0.233)	(0.318)		(0.259)	(0.299)
Treat*FakePost			0.265		
			(0.250)		
Observations	7,638	7,703	4,100	8,721	3,904
R-squared	0.253	0.241	0.183	0.236	0.201
Cohort FE	Yes	Yes	Yes	Yes	Yes
Degree course FE	Yes	Yes	Yes	Yes	Yes

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: The table reports the results of five robustness checks using the (1) specification. In column (1) we exclude the academic year 2007-08, as the Ministerial Decree is implemented. In column (2), we exclude cohort 2009-10, to rule out any effect due to exam reorganization. In column (3), we estimate a DID for pre-reform cohorts using a fake treatment period (being enrolled in 2009-10). In column (4), we exclude English degree courses from the sample, and in column (5), we consider a balanced panel of courses. Standard errors clustered at the degree-course by cohort level.

Table A7: First-year outcomes: Additional robustness checks

	(1)	(2)	(3)	(4)	(5)
(a) University dropout 1st year					
Treat*Post	-0.042**	-0.042	-0.037**	-0.037	-0.057**
	(0.018)	(0.026)	(0.018)	(0.026)	(0.024)
Observations	9,083	9,083	9,083	9,083	9,083
R-squared	0.032	0.032	0.072	0.072	
Method	TWFE	TWFE	TWFE	TWFE	DRDiD
Student-level covariates	No	No	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	
Degree course FE	Yes	Yes	Yes	Yes	
Clustered s.e.	cohort*course	course	cohort*course	course	course
Wild Boot pv:		.1912		.3013	
(b) No. ECTS 1st year					
Treat*Post	11.205***	11.205***	10.486***	10.486***	10.134***
	(1.983)	(2.608)	(1.952)	(3.018)	(2.690)
Observations	9,083	9,083	9,083	9,083	9,083
R-squared	0.135	0.135	0.221	0.221	
Method	TWFE	TWFE	TWFE	TWFE	DRDiD
Student-level covariates	No	No	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	
Degree course FE	Yes	Yes	Yes	Yes	
Clustered s.e.	cohort*course	course	cohort*course	course	course
Wild Boot pv:		.016		.005	
(c) No. exams 1st year					
Treat*Post	1.061***	1.061***	0.974***	0.974***	0.877**
	(0.265)	(0.223)	(0.244)	(0.153)	(0.434)
Observations	9,083	9,083	9,083	9,083	9,083
R-squared	0.233	0.233	0.307	0.307	
Method	TWFE	TWFE	TWFE	TWFE	DRDiD
Student-level covariates	No	No	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	
Degree course FE	Yes	Yes	Yes	Yes	
Clustered s.e.	cohort*course	course	cohort*course	course	course
Wild Boot pv:		.004		.003	

*** p<0.01, ** p<0.05, * p<0.1

Notes: The table reports the DiD estimates of the Treat*Post interaction using clustering at different levels and doubly-robust estimators (DRDiD) following Sant'Anna and Zhao (2020). Estimates using the improved doubly-robust DiD estimator based on inverse probability of tilting and weighted least squares with wild-bootstrap standard errors are reported in column (5).

Table A8: Graduation outcomes: Adjusted DiD estimates

			If graduated		If graduated on time	
	Graduation	On-time graduation	Final grad. mark	*Cum laude*	Final grad. mark	*Cum laude*
	(1)	(2)	(3)	(4)	(5)	(6)
Treat*Post	0.043*	0.073**	2.634***	0.002	2.615***	0.007
	(0.022)	(0.031)	(0.855)	(0.021)	(0.684)	(0.029)
Observations	9,083	9,083	6,641	6,641	4,660	4,660
R-squared	0.134	0.173	0.200	0.048	0.123	0.045
Student-level covariates	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Degree course FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean outcome	.73	.41	92.67	.07	98.39	.12
% effect	5	17	2	2	2	5

*** p<0.01, ** p<0.05, * p<0.1

Notes: The table reports the estimates of the adjusted DiD model using on the graduation outcomes. Standard errors clustered at the degree course by cohort level.

Table A9: Graduation outcomes: Unadjusted event-study DiD estimates

			If graduated		If graduated on time	
	Graduation	On-time graduation	Final grad. mark	*Cum laude*	Final grad. mark	*Cum laude*
	(1)	(2)	(3)	(4)	(5)	(6)
Treat==1 & Yr==2007	-0.019	0.005	0.346	-0.019	1.209	-0.023
	(0.055)	(0.048)	(1.055)	(0.026)	(0.864)	(0.034)
Treat==1 & Yr==2008	-0.014	0.012	0.322	0.010	-0.035	0.021
	(0.039)	(0.041)	(0.885)	(0.015)	(0.718)	(0.022)
Treat==1 & Yr==2010	0.023	0.029	1.308	-0.025	1.779	-0.025
	(0.037)	(0.040)	(1.236)	(0.027)	(1.139)	(0.042)
Treat==1 & Yr==2011	0.069*	0.146***	4.306***	0.021	3.786***	0.023
	(0.041)	(0.040)	(1.183)	(0.021)	(1.066)	(0.027)
Treat==1 & Yr==2012	0.089*	0.175***	5.046***	0.023	3.917***	0.021
	(0.049)	(0.046)	(1.145)	(0.022)	(1.052)	(0.026)
Treat==1 & Yr==2013	0.013	0.122***	6.534***	0.043	5.170***	0.045
	(0.041)	(0.043)	(1.161)	(0.029)	(1.000)	(0.035)
Observations	9,083	9,083	6,641	6,641	4,660	4,660
R-squared	0.044	0.085	0.121	0.037	0.081	0.035
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Degree course FE	Yes	Yes	Yes	Yes	Yes	Yes

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: The table reports the estimates of the unadjusted event-study DiD model on the graduation outcomes. Standard errors clustered at the degree course by cohort level.

Table A10: Graduation outcomes: Adjusted event-study DiD estimates

			If graduated		If graduated on time	
	Graduation	On-time graduation	Final grad. mark	*Cum laude*	Final grad. mark	*Cum laude*
	(1)	(2)	(3)	(4)	(5)	(6)
Treat==1 & Yr==2007	0.006	0.037	0.925	-0.014	1.220*	-0.022
	(0.042)	(0.037)	(0.860)	(0.027)	(0.723)	(0.036)
Treat==1 & Yr==2008	0.001	0.030	0.462	0.010	-0.005	0.020
	(0.032)	(0.033)	(0.695)	(0.015)	(0.622)	(0.023)
Treat==1 & Yr==2010	0.031	0.039	1.037	-0.028	1.619	-0.028
	(0.030)	(0.031)	(1.013)	(0.027)	(1.024)	(0.042)
Treat==1 & Yr==2011	0.057	0.130***	3.603***	0.012	3.425***	0.015
	(0.038)	(0.040)	(1.072)	(0.022)	(0.979)	(0.028)
Treat==1 & Yr==2012	0.086**	0.172***	4.894***	0.022	3.982***	0.023
	(0.042)	(0.040)	(0.942)	(0.023)	(0.947)	(0.027)
Treat==1 & Yr==2013	0.004	0.106***	5.938***	0.037	4.882***	0.039
	(0.034)	(0.035)	(0.970)	(0.030)	(0.908)	(0.036)
Observations	9,083	9,083	6,641	6,641	4,660	4,660
R-squared	0.135	0.174	0.204	0.049	0.126	0.046
Student-level covariates	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Degree course FE	Yes	Yes	Yes	Yes	Yes	Yes

*** p<0.01, ** p<0.05, * p<0.1

Notes: The table reports the estimates of the adjusted event-study DiD model on the graduation outcomes. Standard errors clustered at the degree course by cohort level.

Table A11: GPA at graduation: DiD and event-study DiD estimates

	If graduated		If graduated on time	
	(1)	(2)	(3)	(4)
Treat*Post	0.681***		0.613***	
	(0.211)		(0.179)	
Treat==1 & Yr==2007		0.252		0.509***
		(0.232)		(0.181)
Treat==1 & Yr==2008		0.185		0.148
		(0.216)		(0.172)
Treat==1 & Yr==2010		0.306		0.450*
		(0.270)		(0.240)
Treat==1 & Yr==2011		1.017***		0.977***
		(0.266)		(0.231)
Treat==1 & Yr==2012		1.207***		1.052***
		(0.253)		(0.219)
Treat==1 & Yr==2013		1.565***		1.363***
		(0.255)		(0.220)
Observations	6,641	6,641	4,660	4,660
R-squared	0.129	0.133	0.096	0.100
Cohort FE	Yes	Yes	Yes	Yes
Degree course FE	Yes	Yes	Yes	Yes
Mean outcome	24.17		25.24	
% effect	2		2	

*** p<0.01, ** p<0.05, * p<0.1

Notes: The table reports the estimates of the adjusted event-study DiD model using as outcomes the exam GPA at graduation. Standard errors clustered at the degree course by cohort level.

Table A12: Graduation outcomes: Heterogeneity across high school track — Unconditional DiD estimates

			If graduated		If graduated on time	
	Graduation	On-time graduation	Final grad. mark	*Cum laude*	Final grad. mark	*Cum laude*
	(1)	(2)	(3)	(4)	(5)	(6)
Treat*Post	0.042	0.086**	3.432***	0.012	3.068**	0.027
	(0.033)	(0.040)	(1.277)	(0.028)	(1.284)	(0.045)
Treat*Post*AcademicHS	0.019	-0.001	-0.460	-0.006	-0.332	-0.023
	(0.039)	(0.044)	(1.248)	(0.032)	(1.294)	(0.047)
Observations	9,083	9,083	6,641	6,641	4,660	4,660
R-squared	0.079	0.116	0.139	0.042	0.092	0.041
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Degree course FE	Yes	Yes	Yes	Yes	Yes	Yes
School track	Yes	Yes	Yes	Yes	Yes	Yes

*** p<0.01, ** p<0.05, * p<0.1

Notes: The table reports the estimates of the Treat*Post*AcademicHS (i.e. the DiD term with an academic high school track indicator) interaction in the unadjusted DiD model, to capture potential heterogeneous effects by secondary school track (academic vs. vocational/other). We included in the academic high school track the scientific lyceum and the classical lyceum. Standard errors clustered at the degree course by cohort level.

Table A13: Graduation outcomes: Heterogeneity across high school track — Conditional DiD estimates

			If graduated		If graduated on time	
	Graduation	On-time graduation	Final grad. mark	*Cum laude*	Final grad. mark	*Cum laude*
	(1)	(2)	(3)	(4)	(5)	(6)
Treat*Post	0.041	0.082**	2.956**	0.005	2.887**	0.021
	(0.030)	(0.037)	(1.194)	(0.028)	(1.202)	(0.043)
Treat*Post*AcademicHS	0.007	-0.015	-0.456	-0.003	-0.342	-0.018
	(0.036)	(0.043)	(1.160)	(0.031)	(1.217)	(0.045)
Observations	9,083	9,083	6,641	6,641	4,660	4,660
R-squared	0.137	0.176	0.206	0.051	0.130	0.048
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Degree course FE	Yes	Yes	Yes	Yes	Yes	Yes

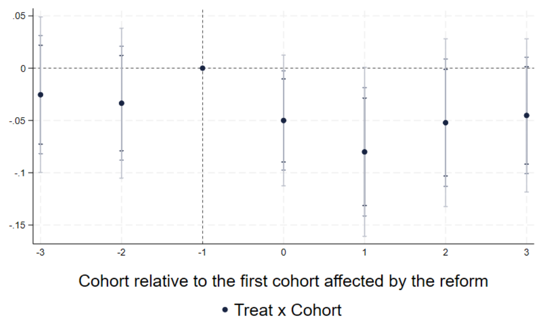
*** p<0.01, ** p<0.05, * p<0.1

Notes: The table reports the estimates of the Treat*Post*AcademicHS (i.e. the DiD term with an academic high school track indicator) interaction in the adjusted DiD model (i.e., controlling for covariates), to capture potential heterogeneous effects by secondary school track (academic vs. vocational/other). We included in the academic high school track the scientific lyceum and the classical lyceum. Standard errors clustered at the degree course by cohort level.

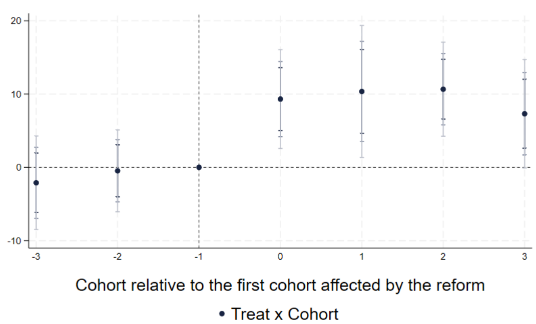
Figures

Figure A1: First-year outcomes: Conditional event-study DiD estimates

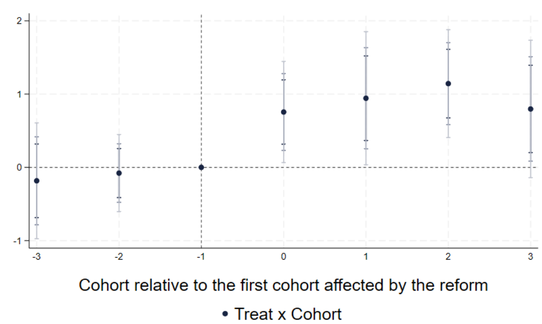
(a) University dropout



(b) No. ECTS 1st year



(c) No. exams 1st year



Notes: The figures display the estimated coefficients associated with the interaction between the $Treat$ indicator and cohort dummies relative to the first cohort affected by the exam reorganization (2010-11), controlling for student characteristics. The baseline is the cohort before the exam reorganization implementation. The dependent variables are the probability of university dropout by the end of the first year of studies (panel a); the number of ECTS accumulated during the first year of studies (panel b); the number of exams accumulated during the first year of studies (panel c). The regression includes cohort and degree courses fixed effects. Standard errors clustered at the degree-course level.