Intensity of Workload per Exam and Academic Outcomes*

Carmen Aina[†] Koray Aktaş[‡] Giorgia Casalone[§]

February, 2022

Abstract

We analyze the effect of a change in the workload allocation across university courses on students' outcomes. In a difference-in-difference setup, we exploit a reform that modified the organization of university degrees in Italy, using an administrative data of a public university. By reducing the number of courses needed to graduate without changing the total workload – expressed in university credits – per degree, the reform mechanically amplifies the average workload per course. Our findings show that an increase in the average workload per course raised the first-year drop-out probability and reduced the probability of graduating for students enrolled in scientific degrees. We document milder results for students enrolled in socio-humanities degrees that were less touched by the reform. Students who graduated after the reform also have higher employment rates. The research reveals that an unintended consequence of the analyzed reform is to operate as a powerful selective mechanism in degrees that do not have admission tests. The selection of students is mainly driven by their ability, and not by other characteristics that should not affect academic outcomes.

JEL codes: I23, I26, I28, J21, J24.
Keywords: number of exams, college drop-out, graduation, workload per exam, higher education.

*Koray Aktaş acknowledges financial support from the Economics Department of the Università del Piemonte Orientale. We thank the participants at various conferences and workshops. We gratefully thank Tommaso Colussi and Diogo G.C. Britto for their insightful comments and suggestions through the development of this paper. This paper has also benefited from conversations with Ingo Isphording, Ulf Zölitz, Anne Boring, Enrico Cantoni, Fabian Waldinger, and Daniela Sonedda.

[†]Department of Economics and Business Studies, Università del Piemonte Orientale, Novara, Italy. E-mail: carmen.aina@uniupo.it.

[‡]Corresponding author. Department of Economics, Management and Statistics, Università degli Studi di Milano-Bicocca, Center for European Studies (CefES), Milan, Italy. E-mail: koray.aktas@unimib.it.

[§]Department of Economics, Università del Piemonte Orientale, Novara, Italy. E-mail: giorgia.casalone@uniupo.it.

I. Introduction

Task management and efficient time allocation across assignments play an essential role in workers' productivity (Coviello et al. 2014; Coviello et al. 2015; Bray et al. 2016). Similarly, when preparing for exams, students must allocate their time efficiently to study a certain amount of material proposed by course teachers. Several studies have investigated the importance of the number of years of schooling at different educational levels (Card 1999; Pischke 2007; Cappellari and Lucifora 2009; Krashinsky 2014; Marcus and Zambre 2018), the amount of coursework during college degrees (Arteaga 2018), and the organization of academic calendars (Bostwick et al. *forthcoming*), but there is no evidence on how the workload allocation across university courses influences the academic performance of college students. In fact, the literature on higher education lacks evidence to suggest the optimal organization of the study time.

In this paper, we document the first evidence in the literature on how the increased workload per course affects the key academic outcomes of college students. For this purpose, we exploit the exogenous variations in the number of courses (or exams) – holding constant the total number of credits students must achieve – across degree programs and academic years.[1] We further investigate the early labor market performance of graduates before and after the change in the organization of the degree programs.

Our identification strategy builds upon the exogenous variations in the number of courses (i.e. exams) induced by a nationwide tertiary education reform.[2] The reform sought to standardize the organization of degree programs across Italy and reduce the time taken to complete a degree. It sets a maximum cap for the total number of courses to be taken for degree completion (for instance, 20 courses to complete a three-year bachelor degree) without making any changes to the number of credits needed for the degree (180 credits). This resulted in a decrease in the number of exams that students sit to achieve a degree that had more courses in their curricula than required by the reform, mostly by unifying two or in some cases three courses (exams) into one. Hence, the intensity of

[1] In the Italian university system, all university courses end with a final exam, as explained in section III. For this reason, in the text we will use alternatively the terms 'course' and 'exam' as substitutes.

[2] Details of the reform and the Italian higher education system will be discussed in section III.

workload per exam has mechanically increased.

We use the administrative dataset of Università del Piemonte Orientale (hereafter UPO), which contains all of the details about the students' enrollment records and performance during their degrees, as well as information about the pre-enrollment characteristics of these students.[3] At UPO, the reform affected the degree programs differently. For example, students enrolled in bachelor of science programs (e.g. biotechnology, biology, mathematics, chemistry, computer science) used to take on average 33 courses to attain their degrees and eight courses in their first-year, whereby these numbers fell to 20 to degree completion and five in the first academic year after the reform, but the number of credits required to graduate (180 credits) did not change, as well as their distribution across years (60 credits per year). Similar but relatively smaller decreases occurred for bachelor degrees in social sciences (e.g. economics, political science, sociology) and humanities (e.g. literature and foreign languages). On the other hand, no changes were made to the number of courses required for bachelor degrees in health care studies organized by the medical school, because the course numbers were already at the level required by the reform.[4]

Therefore, a quasi-experimental environment was set after the requirements of the reform came in force. Such a setup allows us to identify causal effects and makes the use of the UPO dataset appealing. The law passed in 2007, expecting all the universities to comply by the beginning of the 2009–10 academic year. In our setup, science degree programs reorganized their course schedule before the 2009–10 academic year, whereas the social sciences and humanities degrees made the adjustments before the 2008–09 academic year.

In difference-in-difference framework, we consider the three-year degree programs in science, the social sciences, and humanities as treated units, whereas the three-year degree

[3]We provide further information about UPO in Appendix A.

[4]We will refer to these students as students of medical school throughout the paper. Our working sample always only comprises students who enroll three-year bachelor degrees (in both the treatment and control groups).

programs in medical school are considered as the control group.[5] We estimate the impact of sitting unified exams by comparing the differences in the outcomes of the treatment and control groups before and after the reform. While throughout the paper we present results on a variety of academic outcomes, our main focus is placed on the effects on first-year drop-out and degree completion rates. These two academic outcomes of students are the most policy-relevant indicators of higher education efficiency. They are also not subject to endogeneity problems as they are available to all students. Our working sample covers the academic years at enrollment from 2002–03 to 2010–11, which allows us to investigate in depth the validity of parallel trends assumption in both our first-stage and reduced-form regressions.

A priori, the effects of the increased workload per exam are cryptic. Studying for the unified exams can help students to deepen their knowledge on the subject, while taking fewer exams can also mean less stress for students. It can foster non-cognitive skills as students must learn to better manage their time with self-imposing deadlines. On the other hand, evidence shows that students procrastinate and they benefit more from externally-imposed deadlines (Ariely and Wertenbroch 2002). Frequently testing is considered as a motivational equalizer by Tuckman (1998). When students take more exams, they also receive more feedback (De Paola and Scoppa 2011). Evidence also shows that task juggling is less productive than dealing with the tasks subsequently (Coviello et al. 2015). When students take the unified exams, they have to study and process a lot of information for one exam (i.e. task juggling), which can culminate in worsening academic performance. The results of an experiment by Buser and Peter (2012) reveal that individuals who multi-task and who are free to organize their own schedules perform worse than those who are forced to work on the tasks sequentially.

Our results show that after the introduction of unified exams, students in science degrees are on average up to 25 percentage points (p.p.) more likely to drop out at the end of their first year. This number corresponds to almost a 100% increase over the baseline

[5] For convenience, throughout the paper we will present the results obtained from two samples separately. First sample will consist of students in science degree (treatment) and students in medical school (control), and the second sample consists of students in social sciences and humanities degrees (treatment) and students in medical school (control).

drop-out rate of students in science degrees. The latter increase in the first-year drop-out rate mostly – but not entirely – translates into a decrease in the graduation rate (e.g. 22 p.p.). Event-study analysis reveals that parallel trend assumption holds in both our first-stage and reduced-form regressions. The results are also robust to controlling for students' composition through exact matching on the key pre-enrollment characteristics.

However, we do not observe any significant effect on the other outcomes related to the degree completion, namely, time to graduation (measured in months) and final graduation marks. This result reveals that reform did not succeed one of its main goals, namely to reduce the elapsed time to degree.

The findings on the students in social sciences and humanities degrees are somewhat weaker than the results on students in science degrees. Depending on the specification and the sample considered, we find an increase in the first-year drop-out rates ranging from 6.6 to 9 p.p., which translates into a decrease in graduation rates. The relatively smaller effects that we estimate on students in social sciences and humanities degrees can be attributed to the weaker impact of the reform on the number of exams in the curricula of these degrees, as we show in our first-stage regressions.

We further investigate the potential mechanism behind the worsened academic outcomes of students. Similar to the findings on first-year drop-out rates, we estimate a tremendous increase in the probability of not achieving any credit as first-year students. We show that students who fail all of their exams drop out in the end, and such failures mainly drive our main findings. We also show that the average grades of first-year students significantly decrease after the reform. These findings are consistent across the students enrolled in science, social sciences, and humanities degrees.

Our heterogeneity analysis uncovers that high-ability students in science degrees are significantly less affected by the new curricula. In fact, in the first academic year after the reform, these students are more likely to complete their degrees. As for the students in social sciences and humanities degrees, we observe that students who completed an academic high school track are not affected by sitting unified exams.

Finally, we examine the early labor market performance of students who graduated from their degrees after the reform, using the national labor market survey data of *AlmaLaurea*, which is conducted one year after these students completed their studies.[6]

We observe sizable effects on the employment rates of graduate students who were subject to unified exams throughout their degrees. The effect that we estimate for the graduates of science degrees is an average 20 p.p. increase in employment rate, which becomes a 40 p.p. increase when we control for the students' composition by performing exact matching among graduates based on pre-enrollment characteristics. The latter effect takes place for the cohort that is most strongly affected in terms of the first-year drop-out rate, indicating that reduced class size and increased quality of peers after the first year play key roles in explaining the effect on employment. In line with the estimated differential effects across degrees on the academic outcomes, we observe a relatively smaller increase in the employment of students who graduated from social sciences and humanities (e.g. 10-15 p.p. increase in employment).[7] On the other hand, we do not observe any significant effect on monthly net salaries conditional on employment.

The remainder of the paper is organized as follows. In section II, we frame our analysis within the relevant literature. In section III, we provide details about the Italian tertiary education system and the reform that we leverage to identify our estimates. In section IV, we present our dataset, descriptive statistics and descriptive evidence. Section V discusses the identification strategy and outlines the econometric specifications, before section VI presents the results and discusses the underlying mechanisms behind the main findings. In section VII, we show the results from matched samples, heterogeneous analysis, as well as sensitivity analysis. Section VIII shows the findings on labor market outcomes, and finally section IX concludes.

[6] *AlmaLaurea* is an inter-university consortium representing 90% of Italian graduates. It surveys the profile and employment status of the graduates after one, three, and five years.

[7] We also show that the effect on employment is entirely driven by the decrease in the number of graduates searching for a job rather than changes in self-employment or changes in the number of students who are out of the labor force.

II. Related Studies and Contribution to the Literature

Our study contributes to several strands of the literature. First, we provide novel empirical evidence on students' preferences about task management when it comes to preparations for final exams that students must pass to earn the corresponding university credits. Coviello et al. 2015 show that the productivity of judges who task juggle is lower than those who deal with cases subsequently. In line with this evidence, in a higher education context, we show that when students task juggle across subjects for one "big exam" instead of taking multiple final exams, they perform significantly worse, which also severely influences their educational choices. Moreover, De Paola and Scoppa (2011), closely related to our work, show that students who take mid-term exams are more likely to pass the final exam and obtain higher grades compared to the students who take only the final exam.[8] We focus our analysis on the final academic outcomes of students in addition to exam performances, assessing the effects of the intensity of workload per exam rather than the existence of mid-term exams.

Second, we contribute to the studies regarding the organization of academic calendars. While Bostwick et al. (*forthcoming*) show that switching from quarters to semesters negatively affects college students – increasing the first-year drop-out rates and reducing the rate of graduation on time – we provide evidence on a narrower aspect of the organization of academic calendars in terms of the number of final exams that students sit, holding constant the duration of the calendar and the total credits to be achieved by students. Another recent study by Arteaga (2018) provides evidence that a reduction in the amount of coursework required for college degree completion adversely affects human capital accumulation and leads to reduced wages. We add to the literature by documenting that the coursework per exam in a given academic year has also severe effects on students' performance.

Third, our work contributes to the ever-growing literature about the STEM majors in

[8] They conduct an experiment on the students in economics department of a public university based in the south of Italy. They leverage the same higher education reform that we use to identify exogenous shifts in the number of final exams.

tertiary education.[9] STEM majors produce skilled labor and they are crucial for economic growth (Peri et al. 2015). As documented by Chen (2013), the drop-out rate in STEM majors is very high and students in these degrees are more likely to switch their majors. These students are also known to be more over-confident about their skills at the time of college enrollment (Stinebrickner and Stinebrickner 2013). We show that in addition to personal traits of these students, the organization of degree courses in these scientific fields is crucially important and strongly affects academic performances.

Fourth, the suggestive evidence that we provide on the effects of peer composition on employment contributes to the large body of literature studying the peer effects on different education levels. For instance, Brenøe and Zölitz (2020) show that gender composition in high school affects the enrollment of students to STEM subjects.[10] Bostwick and Weinberg (*forthcoming*) study the effects of gender composition in STEM doctoral programs, showing that women who enroll in programs with a higher share of women are more likely to complete their doctoral studies on time. Further evidence is provided by Anelli et al. (2017) on the effects of nationality composition in STEM majors. They show that the increase in the share of foreign classmates reduces the likelihood of graduation of native students. Our findings suggest that exposure to high-skill peers – even if it is only after the first enrollment year – increases the probability of employment right after graduation.

Our study is also linked to a long-standing problem of the low efficiency of higher education institutes in Italy.[11] According to the OECD (2018), despite an increase in educational achievement – mainly for recent cohorts – Italy remains in the bottom of the education distribution for OECD countries and shows a persistent gap with other developed countries.[12] Our findings have policy implications related to these issues of higher education efficiency in Italy. We show that excessively increasing the course

[9]In our setup, the science degree courses include science, technology, and mathematics but not engineering.

[10]On the other hand, Anelli and Peri (2019), show that gender peer composition in high school does not have any impact on the college major choices of students.

[11]See Triventi and Trivellato (2009) for the historical trends in higher education outcomes in Italy.

[12]For instance, in 2015, the fraction of Italy's population aged 25–64 with tertiary education was 18%, while the average in OECD countries was around 35%. For the youngest cohort (aged 25–34), the figures were 25% and 42%, respectively. The small number of graduates is strongly related to persistent high drop-out rates from the Italian university system, despite efforts by the Italian Ministry of Education over the years to increase retention.

workload much can have severe – and negative – impacts on drop-out and graduation rates. However, we also show and discuss that with certain adjustments the organization of course workload can also be used as an ex-post selection procedure that can help students to anticipate their drop-out decisions.

III. Institutional Settings and the Reform

The Italian university system is organized in three cycles according to the Bologna reform process:[13] the first cycle is the three-year bachelor degree (*Laurea*), the second cycle is the two-year master degree (*Laurea Magistrale*), and the third cycle is three-year PhD programs (*Dottorato di Ricerca*). For particular programs (medicine, law, pharmacy, etc.) there is a "single cycle degree" (*Laurea a ciclo unico*) that lasts five or six years. All individuals holding an upper secondary school diploma are eligible to enroll in first-cycle or single-cycle degree programs. A few degree programs have selective admission procedures that can be implemented at the university or national level.

Degree programs are structured in university credits (*Crediti Formativi Universitari* according to the European Credit Transfer System (ECTS)). University credits are the metric used to measure the workload an average student needs to achieve the expected learning outcomes. Each credit approximatively corresponds to 25 hours of student workload, including both lectures and home study hours. To graduate within the legal duration of a degree program, a student should achieve 60 credits per academic year.[14] The first-cycle degree is awarded to students who earn at least 180 credits, while second- and single-cycle degrees are awarded to students who achieve at least 120 and 300 credits (360 for six-year degrees in medicine and dentistry), respectively.

Within this framework, the Ministry Decree 155/2007 (known as *Mussi's Reform*) introduced several changes that potentially affected students' behavior and outcomes.

The main variation was the setting of a cap on the number of courses (and consequently exams) required to earn the degrees. Up to the 2007 reform, universities were allowed to

[13] See Bratti et al. (2008) and Cappellari and Lucifora (2009) for further details about the Bologna Process.

[14] In the Italian university system, for each cycle a minimum period is defined to graduate but not a maximum one.

freely determine the number of courses that composed the study plans, as long as the minimum number of credits needed to earn the degree was 180, 300 or 120 depending on the university degree type. The autonomy allowed to universities in outlying the degree programs had two main consequences. First, degree programs became excessively fragmented, with a huge number of small and barely coordinated courses (exams). Second, the heterogeneity in the setting of the study plans across universities made it difficult for students to move from one university to another, even within the same field of study.

The reform sets a cap of twenty courses (exams) for first-cycle three-year degrees (180 credits), and twelve for second-cycle two-year degrees (120 credits). Degree programs that exceeded this threshold were forced to revise their study plan by combining two or more courses or eliminating courses with few credits. This change was explicitly meant to "improve the efficacy, quality and consistency of the degree programs" by promoting "the cooperation between professors of different courses."[15] Obviously, as the total number of credits to achieve the degree did not change, the unforeseen consequence of this reform has been an increase in the average workload (in terms of credits) for each course. Neither the Ministry Decree nor the reform guidelines provide any explanation about the reason why the cap was setting at this level. Probably the idea was that an average student is able to attend between three and four courses each semester. To the best of our knowledge, there are no indications in the field literature about the most effective organization of the university degree programs, and thus the latter constraint appears to be rather arbitrary.

Universities were expected to apply Mussi's reform within two academic years (2008–09 and 2009–10) following its entry into force. The only constraint was that the whole degree programs offered at the university level in the same "degree class" had to apply the reform simultaneously.[16] Only post-reform freshmen were affected, whereas the rules did not change for pre-reform enrollees.

The timing of the reform coincides with the economic recession in 2008, which can somewhat have an impact on the students' composition at enrollment. According to

[15] Ministry guidelines for the implementation of the new degree classes.

[16] For instance, all of the degree courses belonging to the degree class L35 (mathematics) in the same university had to be reformed in 2008–09 or 2009–10.

the statistics provided by the Ministry of Education (MIUR), the number of first-year undergraduates in Italy decreased about 13.2% from 2003–04 academic year (338,036) to 2009–10 academic year (293,149). Nevertheless, the latter decrease in college enrollment took place rather smoothly than a sharp impact of economic recession.[17] Adamopoulou and Tanzi (2017) provide evidence that the recession in 2008 decreased the college drop-out in Italy. They discuss that the decrease in drop-out probability is an outcome of a decrease in cost of studying during the times in which the youth unemployment rate is high.

To address the concerns about potential changes in the students' composition, we will compare the results obtained from main estimation sample with the ones from a sample that is created through exact matching procedure on the pre-enrollment characteristics of students.

IV. Data and Descriptive Evidence

We use administrative data of the Università del Piemonte Orientale. UPO is a public university based in the Piedmont region of Italy with three campuses in the provinces of Alessandria, Novara and Vercelli. The university offers students a wide range of courses in three-year bachelor programs, two-year master programs and five- or six-year programs. During the academic years that we work with, the teaching activities of the university are organized by seven faculties: science (mathematics, physics, and natural sciences), medicine, pharmacy, law, political science, economics, literature and philosophy. These seven faculties are aggregated into four scientific areas: medicine and pharmacy faculties into medicine (*area sanitaria*), the science faculty into science (*area scientifica*), law, economics, and political science into social sciences (*area sociale*), and literature and philosophy into humanities (*area umanistica*). As previously mentioned, the reform required changes in the curricula and the number of exams at the degree program level. However, at UPO, the reform has taken place at a more aggregate level, namely

[17] The enrollment numbers are 338,036 in 2003–04, 331,893 in 2004–05, 323,930 in 2005–06, 308,185 in 2006–07, 307,586 in 2007–08, 294,932 in 2008–09, 293,149 in 2009–10, 288,876 in 2010–11, 280,119 in 2011-12, 269,988 in 2012–13. These data are publicly available at the website of MIUR: link to the website.

the scientific area level. Degree programs within the same scientific area (science, social sciences, humanities, and medical school) simultaneously adjusted their curricula.

All degrees offered by the medical school (three-year degree programs as well as the six-year medicine degree) perform a mandatory admission test prior to enrollment. For the science degrees, only biotechnology program requires students to take the admission test. On the other hand, all of the social sciences and humanities degrees are open to any students holding an upper secondary school diploma. UPO administrative officials stated that no changes to admission policy occurred during the academic years analyzed in this paper. As we work in a difference-in-difference framework, the latter confirmation reassures that our identification strategy has not been contaminated.

We restrict our sample to bachelor's degree (three-year) students to have homogeneous and comparable treatment and control groups. Students who enroll in master programs and five- or six-year degrees (pharmacy and law, medicine and surgery, respectively) are more likely to be more skilled and motivated than the bachelor's degree students. Moreover, our working sample includes only pure freshmen (i.e. not coming from other degree programs/institutions). Therefore, if a student drops out (graduate) from a degree program and enrolls in another degree program, the information on the second (or more) enrollment is not covered in our final estimation sample.

The timing of the introduction of the reform to the degree programs varies across the scientific areas at UPO, which enables us to have an identification strategy to estimate the effects of the changes in curricula. The degree programs in social sciences and humanities introduced the reform in the 2008–09 academic year, while science degree programs introduced it in 2009–10.[18] Finally, in the 2011–12 academic year, the medical school degree programs applied the changes.

For convenience, we split our working sample into two. The first sample comprises students in the science programs and medical school, while the second sample contains information about students in the social sciences, humanities and medical school programs. The former sample has information for about 7,819 students and the latter for about

[18]These academic years stand for the enrollment years of students. Throughout the paper, when we specify an academic year, it will always refer to the enrollment year of the corresponding cohort.

15,705 students, while both samples cover the 2002–03 to 2010–11 academic years. There is a restriction on the final year of the sample because the medical school applied the changes in curricula and exams at the beginning of 2011–12. Although – as explained above – these changes occurred only as a formality, in the same academic year another nationwide higher education reform restructured the organization of departments and faculties in universities. This could cloud our identification strategy for the estimates after the corresponding academic year.

We present some average statistics about the observable students' characteristics in Table 1. While there are some differences in the compositions of students across different scientific areas (e.g. the share of female students in science degrees is smaller than the share in medical schools) the numbers are fairly close regarding the share of different school tracks and high school final marks. Nevertheless, later in the paper we will fully balance the students' compositions by performing exact matching on these variables.

Our data contain information at the student level. Specifically, we have information on the exact date of enrollment, the enrolled degree, the exact date of exit from the degree with the reason for exit (drop-out or graduation), the date of birth, gender, students' final high school mark, the type of high school diploma, and the province of the high schools that the students graduated from. We also have student-level data on the exams. For each student in a given academic year, we have information concerning which exams were passed, along with the grades earned from these exams. However, it is noteworthy that in exam data, the information about exams only appears if the student passed the corresponding exam. Students must earn a grade of 18 (out of 30) to pass an exam.

IV.I. Descriptive evidence

In this section, we present descriptive evidence on number of exams, which is our first-stage analysis, and the main academic outcomes of interest.

IV.I.I. Number of exams

In Fig. 1, we highlight the average number of exams per scientific area over the years (figures at the top display the corresponding numbers for the first academic year, while those at the bottom relate to the entire degree). These numbers are calculated by only considering the exams taken by students who completed their degree within 45 months, assuming that these students are more likely to have passed their exams "on time."

As shown in Fig. 1, there is a significant reduction in the number of exams for degrees in social sciences and humanities in the 2008–09 academic year, while we observe a similar but even stronger decline in the number of exams for degrees in science at the beginning of the 2009–10 academic year. For example, before the reform, a student of a science degree used to take on average eight exams during her first year, whereas the corresponding number is five exams after the reform. In terms of the number of exams taken throughout science degrees, this falls from an average of 33 exams to 20.

Regarding medical school degrees, we do not see any sort of changes in the number of exams, which is reassuring for our identification strategy as these students define our counterfactual group. The latter also explains the reason why the medical school waited to apply the changes required by the reform until the 2011–12 academic year. Essentially, the number of exams in these programs was already at the required level, and the changes only occurred for them on paper as a formality.

IV.I.II. Academic outcomes

We focus on four main academic outcomes, namely first-year drop-out rates, graduation rates, time to graduation (in months), and final graduation marks of students.[19]

In Fig. 2 and Fig. 3, we present the mean statistics of these variables along with the confidence intervals at the 95% level for our two estimation samples. The upper-left figure in Fig. 2 shows that the first-year drop-out rates of students across science degrees and medical school follow very similar trends over the years until the 2009–10 academic year,

[19] Obviously we can only observe the final marks and time to graduation for the students who complete their degrees, which can raise concern about selection bias. We will address this issue later in the paper when we perform exact matching among graduates.

in which the reform takes place. During the pre-treatment periods, the average drop-out rate of science students is approximately 22%, while this figure is around 14% for the students at medical school. We see a substantial increase with the introduction of the new curricula, whereby specifically the first-year drop-out rate is around 33% for the cohort of the 2009–10 academic year and it reaches almost 50% in the 2010–11 academic year. Although these figures are unconditional averages and we will control for the key observable characteristics of students in our regression analysis, such a substantial increase in drop-outs during the treatment period already reveals the struggles of students with the new curricula.

There is a similar but reverse pattern for the graduation rates (upper-right figure). We see that graduation rates seem to have declined after the introduction of the reform. On the other hand, for the other two figures at the bottom of Fig. 2 concerning time to graduation and graduation marks, we do not see any clear patterns or fractions after the reform, at least descriptively.

Fig. 3 presents the mean statistics of the same outcome variables but for the second estimation sample, which comprises social sciences humanities and medical school.[20] This time, we have a different picture as the changes in the outcomes of social sciences and humanities students in the 2008–09 academic year and onward are rather mild. The fact that the changes in the number of exams in these degrees are much smaller with respect to the changes in the science degrees can explain relatively smaller alterations in the academic outcomes. Nevertheless, we will provide more insights into these outcomes later in the paper, e.g. the impact on heterogeneous groups within treated units.

V. Identification and Empirical Framework

We employ a difference-in-difference approach to identify the effects of exogenous changes in the number of exams on the outcomes of interest. We have two sources of variation, the first being the variation across the timing of the implementation of the changes in curricula. As previously stated, the degree courses in social sciences and humanities

[20]The sample of medical school in Fig. 3 is identical to the sample in Fig. 2.

adjusted their curricula at the beginning of the 2008–09 academic year, and the science degree courses applied the required changes in the 2009–10 academic year. On the other hand, bachelor's degrees in medical school waited until the 2011–12 academic year to execute these changes given that the number of exams in these courses were already at the required level (as shown in Fig. 1). Therefore, we restrict our estimation sample to the academic years between 2002–03 and 2010–11 to use students enrolled in medical school programs as a counterfactual group, which allows us to identify the effect for social sciences and humanities during the academic years from 2008–09 to 2010–11, and for the science degrees during the 2009–10 and 2010–11 periods.

The second source is the information on outcomes of treatment groups before and after the reform. We split the main working sample into two sub-samples. The first one comprises medical school students as a control group and the science program students as the treatment group. In the second sample, social sciences and humanities make up the treatment group and once again the medical school is the control group.

V.I. Baseline econometric specification

The econometric specification that we use is as follows:

$$\begin{aligned} Y_i &= a + \gamma Treated_i + \lambda_1 POST_{1t} + \lambda_2 POST_{2t} + \delta_1 (Treated_i \times POST_{1t}) \\ &+ \delta_2 (Treated_i \times POST_{2t}) + X_i \beta + \epsilon_i; \end{aligned} \tag{1}$$

where Y_i is the outcome of interest, $Treated_i$ is a dummy variable that is equal to 1 if i is a student of the treated scientific area, $POST_{1t}$ is a dummy equal to 1 if the academic year t is the first post-reform year and $POST_{2t}$ is a dummy equal to 1 if the academic year t is the second post-reform year; X_i is the observable characteristics of student i; the parameters of interest that we want to estimate are δ_1 and δ_2, which provide us with the average treatment effects for the first and second academic years, respectively. In order to have more balanced samples across pre- and post-treatment years, we start the estimation samples used in this specification from the 2005–06 academic year. We replicate Equation (1) for each academic outcome of each working sample.

Standard errors are clustered at the degree course and academic year level (Carrieri et al. 2015). In general, the ideal way of clustering standard errors in a difference-in-difference framework is to do so at the "treatment" level. In the cases of an insufficient number of clusters, the evidence shows that wild bootstrapping yields convenient results (Cameron et al. 2008). However, MacKinnon and Webb (2017) provide evidence that if the number of clusters in regressions is less than twelve, wild bootstrapping severely under-rejects the null hypothesis of being equal to zero. In our case, the treatment occurs at the degree program level, we have ten degree courses in first estimation sample and the second sample comprises eleven courses. Nevertheless, we have also reproduced our main results by clustering at the degree course level and performing a wild bootstrap procedure. These results are in line with described in what follows and they are available in Appendix B.[21]

V.II. Event-study specification

In addition to our baseline model, Equation (1), we also set up an event-study specification to check whether the common trends assumption of the difference-in-difference approach is satisfied in our estimation samples. For this purpose, we simply interact the year dummies with the treated scientific field dummies. We choose the year before the intervention as a baseline, which is the 2008–09 academic year for the science and medical school sample, and 2007–08 for the social sciences, humanities and medical school sample.

We set up the following model as an event-study specification:

$$Y_i = \alpha + \gamma Treated_i + \sum_{k=2002}^{2010} \lambda_k P_k + \sum_{k=2002}^{2010} \delta_k (Treated_i \times P_k) + X_i \beta + \epsilon_i; \qquad (2)$$

[21] For some outcomes of interest, we lose power in our treatment effect estimates and they become borderline insignificant. As discussed above, this is due to an extremely small number of clusters in the analysis. Furthermore, in a context of an academic environment, the curricula and assignment of lecturers is updated at the beginning of each academic year, so two-way clustering at the degree program and academic year levels is much less restrictive than clustering at – say – state and calendar year levels (e.g. to evaluate a labor market reform occurs at the state level).

where $k = 2002, 2003, ..2010$, P_k are the dummy variables, which are equal to 1 in year k, $Treated_i$ is equal to 1 if student i is in the treated scientific area, X_i is the observable characteristics of student i, the coefficient estimates of interaction terms, δ_k, are the parameters of interest in the model. As in our main specification, we cluster the standard errors at the degree program and academic year levels.

VI. Results

Taking advantage of the fact that we have outlined the econometric specifications that will be used in the paper, we first present the estimation results in Fig. 4 on the number of exams that students take.[22] These results are obtained from Equation (2), and they provide more formal information – with respect to descriptive evidence discussed in subsection IV.I – on how the reform affected the number of exams that the graduates have taken during their degree courses.[23] In the left column of the Fig. 4, we see that after the introduction of the reform in the 2009–10 academic year students of science degrees sit about four exams fewer during their first year (upper-left figure) and this number is around fourteen for the entire degree program (bottom-left figure) with respect to the number of exams that the students of medical degrees sit.

Looking at the right column of Fig. 4, the relative decreases in the number of exams that the students of social sciences and humanities degrees sit are about two exams for the first year and six exams for the entire degree courses. These numbers are about a half of what we observe for the students of science degrees. The reason for the differential effects of the reform on the number of exams across scientific fields is the number of exams that these degrees had prior to reform, as we already showed in Fig. 1, which is one of the main reasons why we split our estimation sample into two.

It is worth noting that the number of exams in medical degrees (control group) is not affected by the reform (see Fig. 1) and all of the relative differences in Fig. 4 are driven by the effect of the reform on the number of exams of our treatment groups, namely science

[22] These results are obtained from regressions without using any control variables on students' characteristics.

[23] As we stated in subsection IV.I, to calculate the number of exams we look at the students who completed their degrees within 45 months.

degrees, and social sciences and humanities degrees.

In the rest of this section, we report the estimation results obtained from Equations (1) and (2) for the first-year and degree outcomes, separately.

VI.I. Effect on the students of science degrees

We present the results obtained from Equation (1) on the average effects of the workload increase per exam on the academic outcomes of students of science degrees in Table 2.

Columns (1), (3), (5), and (7) report findings without the control variables included, while columns (2), (4), (6), and (8) show the results after the inclusion of the control variables on students' observable characteristics. Although the inclusion of the control variables slightly reduces the size of the estimated effects, we observe statistically significant estimates in both cases. After controlling for the key observables of students, we see still an increase of 13 p.p. in the drop-out probability of first-year students in the 2009–10 academic year, while this figure is about 25 p.p. in the 2010–11 academic year (column (2)). Considering the baseline of the first-year drop-out rate of students in science degrees – which is about 23% – the estimated effects on average correspond to approximately a 100% increase over the baseline. This large effect is one of the most striking findings of our study as the first-year drop-outs are considered an important indicator of higher education efficiency.

In columns (3)–(4), the results on the graduation rates are presented. In line with the effects on the first-year drop-out probability, we see negative effects on the graduation rates. For this outcome, we place a cap on the duration of graduation, which is set to 67 months from enrollment. The latter duration for the graduation is the longest observed in our data for the cohort of 2010–11. This restriction ensures that the effects on graduation are not driven by some mechanical relationship between the enrollment years and the year when our data is constructed (September 2016).[24]

An interesting result is the difference in the effects across first-year drop-out and graduation rates for the 2009–10 cohort ($Treated * Post_{1t}$). While we observe a significant

[24] We have produced the same results without this cap and they are perfectly in line with what is reported in this section.

12 p.p. relative increase in the former, the effect on the latter is a 5 p.p. of the relative decrease and statistically not significant, which suggests that students who decided to drop-out at the end of their first year would have dropped out anyway in the subsequent years.[25] This suggestive evidence indicates that increased intensity of exams in the first academic year can actually be used as an ex-post selection procedure to anticipate the drop-out decisions that would eventually take place. Nevertheless, we do not observe the differential effect between the first-year drop-out and graduation rates for the cohort of 2010–11.

As for the control variables, apart from the statistically insignificant coefficient estimate for gender, all other coefficient estimates are in line with the initial expectations;[26] for example, the higher the students' high school graduation marks, the less likely they drop out at the end of their first year. Moreover, students who graduated from high schools with a scientific curriculum are less likely to drop out compared to students who completed high schools with other curricula.

On the other hand, we do not observe statistically significant changes in the time needed to complete a degree course, nor the final graduation marks. The result on the time to degree is controversial as one of the main reasons to place a cap on the number exams was to help students to complete their degrees *on time*. As for the final marks, the results suggest that having more concentrated exams throughout degrees do not directly affect the human capital accumulation. Nevertheless, these two outcomes can only be observed for students who completed their studies successfully. Since we already know that our treatment has a sizable and significant impact on the first-year drop-out rates, the coefficient estimates on these two outcomes conditional on graduation might be to some extent biased due to changes in the compositions of graduates after the reform. We will address this concern of selection bias when we present our exact matching strategy in the next section.

[25] This difference in effects between the drop-out and graduation rate is not driven by the slightly different sample selection for the graduation outcome. We have replicated our model on drop-outs for the sample used in estimating the effect on graduation, and the results are the same with what is reported in columns (1)–(2) of Table 2.

[26] Literature shows that female students are more likely to drop out of STEM subjects. See, among others, Griffith (2010); Chen (2013); Isphording et al. (2019).

We present the results from our event-study specification outlined in Equation (2) on the first-year drop-out rate and degree outcomes in Fig. 5 and in Fig. 6, respectively. We see in these figures that the parallel trend assumption holds in our data, which is crucial for our identification strategy. Not only the coefficient estimates are not statistically significant during the pre-treatment period, but they are also very close to zero. This also helps to reassure that the results presented in Table 2 are not affected by the choice of different baselines.

VI.II. Effects on the students of social sciences and humanities

In Table 3, we present the results on the same outcomes for our second estimation sample in which the treatment group comprises students of social sciences and humanities degrees, while the control group is the same of the first estimation sample, namely students of medical school. The results for this sample is rather different than the previous one. When we look at the effects on the first-year drop-out rate, the estimated effect for the 2008–09 academic year ($Treated * Post_{1t}$) is 6.6 p.p. (column (1)) without controlling for the students' characteristics. However, the latter result drops to 5 p.p. and it is no longer statistically significant once we include the control variables (column (2)).

Observing smaller and statistically weaker effects for the students of social sciences and humanities is in line with earlier discussion in this section about the intensity of treatment in these degrees. The degree courses of these fields were subject to fewer changes in the number of exams than the degree courses in science, hence explaining the weaker impact. Another potential explanation for such weak impact is the relatively less challenging curricula of social sciences and humanities degrees compared to the science degrees.

Columns (3)–(8) of Table 3 reports the findings on degree outcomes of students enrolled in social sciences and humanities programs. The results are small in size and always statistically insignificant. This unsurprising considering the mild and borderline significant effects on the first-year outcomes of these students. Taken together, we can conclude that a lower intensity of treatment seems to be an important factor that can account for the results on the academic outcomes of students in social sciences and humanities degree programs that are between weak and non-existent, as opposed to the results for the science degrees.

While on average we do not observe any effect on the academic outcomes of these students, we will investigate the heterogeneous effects on different sub-groups later in the paper.

In Fig. 7 and Fig. 8, we also highlight the results from event-study specification that includes the estimates on the third academic year after the reform – 2010–11 – in addition to what we have presented in Table 3 since our dataset allows us to do so. However, as can be seen from Fig. 8, in the third year of the reform the effects essentially level off in every outcome, which might be caused by some departmental adjustments after observing worsening in students' performance, such as the increased number of students who fail their exams in their first year. This is why we have focused on the first two years of the reform as the changes are rather exogenous to both students and the academic staff.

VI.III. Underlying mechanisms

To better understand the potential channels of the increase in first-year drop-outs after students sit more concentrated exams, we look at some other relevant variables. We create two outcome variables to complement the findings on the first-year drop-outs: the probability of not achieving any credit during the first academic year when students are enrolled, and probability of dropping out at the end of the academic year of enrollment without achieving any credit. We will also investigate the grades of first-year students in this section.

As explained in section IV, our data is created by different sources provided by the UPO administrative office. In the data that provides the information on exams, grades, and credits, we can only observe students if they pass at least one exam. Therefore, we create a dummy variable that takes the value of 1 if a student has a record in the freshmen enrollment data but not in the exam data of the same year, otherwise it is equal to 0. If this variable is equal to 1, it means that student did not pass any exams in her first-year, i.e. did not achieve any credit.[27]

For the second variable, we create a dummy which takes the value of 1 if a student drops out at the end of her first year without achieving any credit.

Fig. 9 (for the first estimation sample) and Fig. 10 (for the second estimation sample) highlight the results obtained from Equation (2) (in the right columns) along with the averages (in the left columns) of these outcomes over the academic years.

In Fig. 9, we see that the probability of not achieving any credit increases by 17 p.p. in 2009–10 and 30 p.p. in 2010–11. Similar but slightly smaller effects are estimated on the first-year drop-out probability without achieving any credit. These findings are perfectly in line with the effect on the first-year drop-outs previously discussed. The findings reveal that the effects presented in the paper are directly driven by the exam-related issues that these students faced after the reform.

We also estimate significant increases in the probability of not achieving credits in social sciences and humanities degrees by 10 p.p. (Fig. 10). The fact that we do not observe any significant effect on the first-year drop-out rates of these students indicate that even though these students fail more exams after the reform, they did not drop out from their degrees.

The final outcome that we investigate regarding the first-year students is exam grades. Grades are represented by a left-censored variable that varies from 18 to 30, as we can only observe a student's grade in the data if she passes the corresponding exam, namely her final grade is either equal to or greater than 18. Similar to the empirical strategy in De Paola and Scoppa (2011), we adjust our event-study specification for this left-censored variable and use a Tobit model.[28]

In Fig. 11, we highlight the estimated marginal effects of the interaction terms on the expected values of grades conditional on being uncensored. Fig. 11 at the top shows the results for the students of science degrees and at the bottom the results for social sciences

[27]However, we cannot possibly know whether these students actually sat on an exam during their first year

[28]We can illustrate the specification of the outcome variable in our Tobit model as follows: $Grade_i^* = \alpha + \gamma Treated_i + \sum_{k=2002}^{2010} \lambda_k P_k + \sum_{k=2002}^{2010} \delta_k (Treated_i \times P_k) + X_i \beta + \epsilon_i$, where $Grade_i^*$ is the latent outcome; the observed grades are $Grade_i = max(0, Grade_i^*)$.

and humanities. We see that students in science degrees obtain lower grades by about 2 points in the 2010–11 academic year, and we also observe a similar result for the second sample for the 2008–09 academic year. We are aware that these results might be to some extent biased as the significant effects that we observe take places during the academic years in which the probability of failing exams is higher. Nevertheless, the findings on grades are in line with the evidence in the literature (see, e.g., De Paola and Scoppa 2011), while the grades are not at the center of our study as we rather focus on the final academic outcomes.

VII. Exact Matching

In addition to the common trend assumption – which has been shown to hold in our estimation samples – the difference-in-difference framework requires stable composition across treatment and control groups when working with repeated cross-section data. In this section, we create estimation samples by performing exact matching on the observable student characteristics. We then highlight the trends in student composition over the years and determine whether there are any significant differences in those trends to see the quality of the matching procedure. Finally, we present the results obtained from the matched samples.

VII.I. Matching procedure

We perform a two-stage matching process. First, in a given academic year (for years both before and after the reform), we exact match treated students with the counterfactual group according to gender, age and the type of high school diploma. For these matched students, we calculate the differences in their high school graduation marks, whereby we keep students in the control group if the difference between their final marks and their counterpart's final mark is not greater or lower than one unit.[29]

[29] A matching procedure on high school marks can be also executed as an exact match, but this leads us to an extremely small sample. Therefore, we perform matching by accepting a small gap between the final marks. High school final marks range between 60–100, so allowing a maximum of a one unit gap for the differences in these marks is reasonable.

In the last stage, we randomly selected two students from the pool of matched control students for each treated student. Considering that our main estimation samples are modest in terms of sample size, in some cases there is no proper match for the treated students, in which case those treated students are excluded from sample. Furthermore, in some cases, one control student is a match for multiple treated students, but this is not problematic in our setup because we work with repeated cross-section data in which we observe each student only for one year. We also allow the observations of control students to repeat based on how many times they are matched with a treated student (i.e. matching with replacement). This is broadly equivalent to weighted propensity score matching and increases the precision of the matching procedure.[30]

Ultimately, our first matched sample comprises 1,709 students from science programs (treated) and 1,480 students (control) from medical school, while the second matched sample contains 3,797 students (treated) from social sciences and humanities programs and 2,428 students (control) from medical school.

We present the trends in the student composition – by age, gender, high school diploma, and high school final marks – in Fig. 12 and Fig. 13 for the first and second estimation samples, respectively. We observe that our matching procedure successfully balances the composition of students at the enrollment.

VII.II. Results on drop-out and graduation rates

Fig. 14 highlights the results on first-year drop-out (the figure on top) and graduation (the figure in bottom) rates for the students of science degrees. The results are mostly in line with the previous findings, with the exception being the size of the effects in the 2009–10 academic year. In Fig. 14, we see an increase of almost 17 p.p. (with a p-value of .072) in the first-year dropout rate, and it entirely translates into a decrease in graduation rate.

As for the results on the second sample, presented in Fig. 15, social sciences and humanities vs. medical school, we see borderline significant estimates on the first-year drop-out rate of students for both academic years, while these coefficients were previously

[30] This final adjustment in creating the matched samples does not alter the results in a significant way. The results obtained from matched samples without replacement are available if requested.

insignificant. Nevertheless, despite the significant negative effect on the first-year drop-out rates, we still do not observe any impact on the graduation rates of these students. As previously discussed, the two potential reasons for such differences in the findings across the two estimation samples are the intensity in the reduction of the number of exams induced by the reform and the arguably less challenging curricula of social sciences and humanities compared to science degrees' curricula.

Overall, estimating the same treatment effects even after balancing the data based on the key ex-ante academic performance indicators of students is quite reassuring for our empirical analysis, indicating that the main findings of this study on first-year outcomes are not driven by the compositional changes induced by the treatment.

VII.III. Matching conditional on graduation

We take a slightly different approach for the graduation outcomes, namely time to graduation and graduation mark. While the matching procedure remains exactly the same, we change its timing. Instead of matching students at the enrollment, we match them conditional on graduation, given that we observe significantly large effects on first-year outcomes, which can also alter the composition of *survivors* from the new curricula.[31]

We present the results from these matched samples in Fig. 16 and Fig. 17 for the first and the second estimation samples, respectively. In line with the findings obtained from our main sample, we do not see any significant effect on students' time to graduation and graduation marks for either of the working samples.

VII.IV. Heterogeneous effects

We now investigate whether there are heterogeneous effects coming from the treatment on several sub-groups. Using our matched samples, we examine the results by gender, academic track high school diploma, ability, and family income. The analysis in this section will help us to further understand the underlying mechanisms that affected students' academic outcomes from students characteristic point of view, and it will complement the

[31] We will show the heterogeneous effects of our treatment in subsection VII.IV.

discussion on the mechanisms in subsection VI.III.

To do so we expand our baseline econometric model outlined in Equation (1) and introduce triple interaction terms.

$$
\begin{aligned}
Y_i &= a + \gamma Treated_i + \omega Het_h + \psi(Treated_i \times Het_h) \\
&+ \lambda_1 POST_{1t} + \lambda_2 POST_{2t} + \delta_1(Treated_i \times POST_{1t}) \\
&+ \delta_2(Treated_i \times POST_{2t}) + \Omega_1(Het_h \times POST_{1t}) + \Omega_2(Het_h \times POST_{2t}) \\
&+ \pi_1(Treated_i \times Het_h \times POST_{1t}) + \pi_2(Treated_i \times Het_h \times POST_{2t}) \\
&+ X_i\beta + \epsilon_i; \qquad (3)
\end{aligned}
$$

where, Het_h is a dummy that represents the sub-group of interest (female, academic track high school, high-ability, high-income family). The definitions of other variables in Equation (3) remains as in Equation (1). The parameters that provide us the heterogeneous treatment effects are π_1 and π_2. We replicate the analysis for each sub-group.[32]

Our focus is on the first-year drop-outs and graduations. In the end, these are the most important academic outcomes and they are free of any concerns regarding selection bias since these two outcomes are available for any student enrolled in any academic year. Table 4 presents the mean statistics in matched samples that will be used in this section. We see that the standardized differences are quite small both before and after the reform – Imbens and Rubin (2015) suggest the corresponding figure should be smaller than .20 – for the variables related to the compositions of treatment and control groups. The differences in levels of the outcomes – first-year drop-out and graduation rates – will be taken care of by our empirical strategy as we employ a difference-in-difference approach.

In Table 5, we report the results for the students of science degree courses. We do not observe any significant result for female students (columns (1) and (2)), suggesting that

[32] For the estimation sample of science degree courses, π_1 and π_2 show the results for the 2009 and 2010 enrollment years, respectively. For the sample of social sciences and humanities, π_1 and π_2 show the results for the enrollment years 2008 and 2009.

they suffer from increased workload per exam as much as male students.[33] Considering that there is a growing body of literature on the gender gap in STEM (science, technology, engineering, mathematics) subjects,[34] our findings on female students performance carry important implications in understanding their academic performance in STEM subjects.

In column (3), for the 2010–11 academic year, we see that students who completed academic track high schools are 12 p.p. less likely to drop out at the end of their enrollment year with respect to treated students who have a high school diploma from technical, professional and other school tracks. However, in column (4) we do not observe any significant heterogeneous effects for these students in terms of their graduation rates, suggesting that these students drop out in the subsequent academic years.

We use the high school final marks as a proxy for ability and students with a final mark above 85 – which correspond to the top 25% of the distribution – who are considered high-ability students. Recall that these results are obtained from a matched sample in which high school marks and tracks are balanced over the years across treatment and control groups, while we control for the school tracks in the regression. In columns (5) and (6), we present the results for the high-ability students. In line with theoretical expectations, these students are relatively less affected by sitting unified exams. In fact, the high-ability students enrolled in 2009 have a higher likelihood of completing their degree by about 20 p.p.[35]

We also investigate whether there are any differential effects for the students of high-income families. For this purpose, we consider the tuition fee that students are supposed to pay. The tuition fees in Italy are based on the (equivalent) students' family income. Tuition fee amounts are set by each university, following national regulations. The lower the family income, the lower the tuition fee that students should pay, up to a complete exemption. For students from very low-income families, scholarships are also available (e.g. food stamps, free accommodation, etc.). We consider the top 25% of the log tuition

[33] This finding is in line with the evidence documented by Buser and Peter (2012) in which they show that women do not perform better than men when multitasking.

[34] See, e.g., Kahn and Ginther (2018).

[35] The estimated heterogeneous effect on graduation rates for this cohort is 0.245, while the estimated coefficient associated with the baseline, $Treated * Post_1$, is -0.055 and it is not statistically different than 0.

fee distribution as the students of high-income families. These students might have access to extra resources not only during their degrees but also ex-ante college enrollment that can help them in developing more non-cognitive skills. However, as we see in columns (7) and (8), we do not find any significant differential effects for these students.

Table 6 presents the results for the students of social sciences and humanities. For the cohort of the 2009 enrollment year, we see an average 13 p.p. relative decrease in the graduation probability of female students (column (2)).

We find significant effects on the outcomes of students who have academic track high school diplomas. These students are less likely to drop out at the end of first year and more likely to graduate with respect to the students from other school tracks. This finding also reveals that students in social sciences and humanities with a high school diploma from technical and professional schools are indeed affected by the reform. Their drop-out rate significantly increased (by 9.5 p.p. in 2008 and 8 p.p. in 2009), while the graduation rates decreased even more so (18 p.p. in 2008 and 8.6 p.p. in 2009).

We do not observe any statistically significant finding for the students of high-income families (columns (7) and (8)). An interesting result is the negative effect on the graduation rate of high-ability students (column (6)). It is not easy to explain such a finding, which is so at odds with the theoretical expectation. Perhaps, initially these highly skilled students chose social sciences to complete their degrees rather quickly, but the new demanding curricula forced them to switch majors that are more appropriate for their skill sets.

VII.V. Sensitivity analysis

In this section, we present further robustness tests to check the validity of our findings. These sensitivity checks will be carried out by using the matched samples explained in section VII.

VII.V.I. Biotechnology vs. Medical School

One might be concerned about the differences in admission procedures between the degrees in our treatment and control groups. As we explained in section IV, the degrees organized by the medical school (i.e. our control group) perform ex-ante admission selection tests, while the degrees in our treatment groups do not carry out such a selection procedure with the exception of biotechnology degrees. Even though these admission policies of degree programs do not change over time and we control for the time-invariant heterogeneity in our empirical setup, we provide this further robustness test by exploiting the unique situation of biotechnology degree in our estimation sample.

At UPO, the biotechnology degree is offered by the medical school, although officially – based on the definition decided by the Italian Ministry of Education – biotechnology degrees are labeled as a science degree and they are subject to any changes in the regulations regarding science degrees. Therefore, students in biotechnology degrees represent a perfect treatment group in terms of comparison with the students in our control group, namely the medical school. We create our working sample for this sensitivity test by performing the matching procedure – explained in the previous section – on students in the biotechnology degree and our usual control group.

We replicate our event-study specification for the first-year drop-out and graduation rates. The results are presented in Fig. 18. The results are striking as we observe an average increase of 45 p.p. in the first-year drop-out rates, which completely translates into an effect on graduation rates. This finding shows that the existence of the admission test in the degrees in our control group does not generate our results.

VIII. Labor Market Outcomes

In this section, we focus on the early labor market outcomes of graduates. We merge our administrative dataset with the *AlmaLaurea* labor market survey, which covers almost all graduates in Italian universities.[36] The merging is based on the fiscal codes of students. Students are contacted for the survey after one year from their graduation. We investigate two labor market outcomes of students, namely employment and net monthly labor

[36] *AlmaLaurea* is an inter-university consortium supported and funded by the member Universities, the Ministry of University and Research, and companies and bodies that use the services offered. We were only allowed to have access to the information of the graduates of UPO.

income, which are the most consistently-observed outcomes in the data.[37]

We find a very high portion of these students in the labor market dataset, from 82% to 92% of the graduates depending on the degree course. However, the analysis on labor market outcomes is limited to the students who are either employed or searching for a job (i.e. students who are out of labor force are excluded). Hence, the sample coverage decreases slightly further. Ultimately, the final sample comprises 636 graduates of science degrees, 2,262 graduates of medical school over the academic years from 2002–03 to 2010–11. For the second estimation sample spanning the same academic years, there are 1,774 graduates of social sciences and humanities and 2,456 graduates from medical school.

In addition to the analysis with the main sample, we will also present results on wages and employment for the matched sample among graduates. As previously discussed, our treatment heterogeneously changes the composition of students after the first year of enrollment. Therefore, to balance the sample in terms of the students composition among graduates, we perform matching conditional on graduation. This creates a comparable sample by netting out the initial effects of our treatment on the first-year drop-out rates.

The matched sample comprises students of science degrees and medical school. We find 804 students (311 science, 492 medical school) in the labor market dataset. For our second estimation sample, the corresponding number is 1,789 (869 social science and humanity, 920 medical school).

Using the econometric specification explained in Equation (2), we present results on wages and employment for the first and second estimation sample in Fig. 19 and Fig. 20, respectively. In each figure, we present the results of full sample in the left columns and the results of matched samples in the right columns.

We see a significant relative increase in the employment probability of the science degree graduates who were subject to new curricula at their enrollment in the 2010–11 academic year. The estimated effect is 20 p.p. in the full sample, while it is 40 p.p. once we exact match the treated and control units. We know from the earlier analysis of

[37] The net monthly salary is reported in the data at intervals of 250 Euro, starting from 0-250 to 3,000+ Euro. We consider the mid points of these intervals as our monthly income measure (e.g. if it is 1,000–1,250 Euro, we consider it as 1,125.5 Euro).

heterogeneous effects of the treatment that high-skill students are more likely to survive the more challenging exam program and that almost 50% of the students who enrolled in 2010–11 drop out at the end of their first year. Such drastic changes in the class sizes and skill distribution of students – which means that the survivors are exposed to more high-skill peers for the rest of their degrees – can help these students to develop non-cognitive skills during their studies. Although our data do not hold any information that can be used to reveal such a mechanism affecting the employment of these students, the evidence on peer effects in the literature is in line with such intuition (Epple and Romano 2011).

It is worth noting that in our labor market survey data we can identify several important insights into the students' employment or unemployment status. Students can describe their employment status as follows: employed, employed and enrolled in masters' degree, enrolled in masters' degree, not searching for a job, and searching for a job. Our outcome measure is created as being employed or employed while enrolled in masters' degree (dummy equal to 1) versus searching for a job (dummy equal to 0). We confirm that there is no change in the probability of not searching for a job or being enrolled in masters' degree (without being employed).[38] The sizable effect that we estimate on employment is driven by the shifts from the status of searching for a job to being employed.

On the other hand, we do not estimate statistically significant effects on wages. Despite the slightly increase in wages from the matched sample, it is not significant.

In line with the effect that we estimate on the employment of science degree students, we also estimate a significant increase in the employment of social sciences and humanities graduates after the reform. Interestingly, we also observe an increase in monthly salaries of these students in the full sample. Despite the fact that we lose the statistical significance in these findings once we perform exact matching, it remains important suggestive evidence.

IX. Conclusions

In a world of growing complexity of tasks, it has become relevant to investigate whether time allocation across assignments is efficient, as it entails the level of productivity

[38]These results are available upon request.

attained. Nevertheless, this link has mainly been explored for workers, while essentially not being analyzed for university students. To fulfill the gap, this study documents the first evidence about the relationship between the intensity of workload per course and the key academic outcomes of university students by exploiting an exogenous variation in the number of courses induced by a national tertiary education reform. Accordingly, we investigate the impact of sitting unified exams (i.e. increased workload per course) by comparing the differences in the college outcomes of treatment (science and social sciences and the humanities degrees) and control (medical school) groups before and after the reform.

Our findings are stark. For students of science degrees, we find that the introduction of the reform increased the first-year drop-out probability by 13 p.p. and 25 p.p. during the first and second year of its implementation, respectively. By contrast, for students of social sciences and humanities this effect is milder, and once the control variables are included any effect disappears. This evidence is also corroborated by further investigation of the potential mechanisms at work, such as early drop-out without any credit and first-year grades. Concerning the graduation outcome, for science students we show a negative effect, albeit only in the second year of application of the reform. Likewise, for social sciences and humanities the effect is small in size and statistically insignificant. Overall, these results confirm that the higher intensity of the treatment applied to scientific degrees drives the worse academic performances of the corresponding subset of students.

In addition, the results obtained from event-study specifications reassure about the existence of parallel trends in the analyzed outcomes prior to the introduction of the new curricula. By performing exact matching on the observable student characteristics, we also show that these results are not driven by changes in the composition of students. To test potential heterogeneous effects on drop-out and graduation outcomes, we also run estimates on the matched sample by gender, academic track, ability and family income. The findings for science degrees reveal no gender and family income differences in the effect of increasing workload per exam, but an advantage of 12 p.p. for the academic high school track in terms of persistence and 20 p.p. for high-ability students (i.e. with a high school grade higher than 85/100) in terms of graduation. Indeed, regarding students of social sciences and humanities, we find a lower first-year drop-out probability for those with an academic high school, but no effects for high-ability students.

Finally, looking at the early labor market outcomes of graduates, we find that graduates exposed to the reform have better labor market prospects as their employment probability is significantly greater. Two potential mechanisms can account for these effects. First, students in our treatment group have to organize their time to study for the unified exams, which can lead to them developing or improving certain skills associated with non-cognitive abilities (e.g. perseverance, planning).[39] The second channel is the increased average peer quality in degree courses, as shown in our heterogeneity analysis, namely that high-skill students are less expected to be affected by the unified exams and more likely to survive from the first year and complete their degrees. This indicates that the peer quality among *survivors* of the first academic year increases. To corroborate this explanation, we demonstrate that the employment rate is higher for the cohorts with a higher first-year drop-out rate after the reform, suggesting that the effect of changes in peer composition is mostly likely the main driving mechanism.

In conclusion, our results indicate that a sizable change in the workload per exam generates a negative impact on the academic performance of undergraduate students, especially enhancing their probability of first-year drop-out. This effect is greater the more that the workload increases. The policy implications are then important. The research tells us that an unintended consequence of the analyzed reform is to operate as a powerful selective mechanism in degrees that do not have admission tests. The selection of students is mainly driven by their ability, as measured by their high school final grade. Other students characteristics that should not affect their academic outcomes (i.e. gender or family income) do not play a significant role in shaping the effect of the reform. This selection improves students' composition and consequently enhances the quality level of the peers. Moreover, it allows students to develop non-cognitive skills in terms of time management and task organization, which seem to facilitate transition to the labor market.

[39] Heckman et al. (2006) provide evidence that non-cognitive skills are significantly associated with the labor market outcomes of individuals.

As a result, the new organization of the degree programs can increase the efficiency of the university system, facilitate the transition to the labor market, and simultaneously avoid wasting the time and financial resources of mismatched students.

References

Adamopoulou, E. and G. M. Tanzi (2017). Academic drop-out and the great recession. *Journal of Human Capital 11*(1), 35–71.

Anelli, M. and G. Peri (2019). The effects of high school peers' gender on college major, college performance and income. *The Economic Journal 129*(618), 553–602.

Anelli, M., K. Shih, and K. Williams (2017). Foreign peer effects and STEM major choice. *CESifo Working Paper Series*.

Ariely, D. and K. Wertenbroch (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological science 13*(3), 219–224.

Arteaga, C. (2018). The effect of human capital on earnings: Evidence from a reform at Colombia's top university. *Journal of Public Economics 157*, 212–225.

Bostwick, V., S. Fischer, and M. Lang (Forthcoming). Semesters or Quarters? The Effect of the Academic Calendar on Postsecondary Student Outcomes. *American Economic Journal: Economic Policy*.

Bostwick, V. and B. Weinberg (Forthcoming). Nevertheless she persisted? Gender peer effects in doctoral STEM programs. *Journal of Labor Economics*.

Bratti, M., D. Checchi, and G. De Blasio (2008). Does the expansion of higher education increase the equality of educational opportunities? Evidence from Italy. *Evolution of Labour Market Inequalities 22*(s1), 53–88.

Bray, R. L., D. Coviello, A. Ichino, and N. Persico (2016). Multitasking, multiarmed bandits, and the italian judiciary. *Manufacturing & Service Operations Management 18*(4), 545–558.

Brenøe, A. A. and U. Zölitz (2020). Exposure to more female peers widens the gender gap in stem participation. *Journal of Labor Economics 38*(4), 1009–1054.

Buser, T. and N. Peter (2012). Multitasking. *Experimental Economics 15*(4), 641–655.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics 90*(3), 414–427.

Cappellari, L. and C. Lucifora (2009). The "Bologna Process" and college enrollment decisions. *Labour Economics 16*(6), 638–647.

Card, D. (1999). The causal effect of education on earnings. In *Handbook of labor economics*, Volume 3, pp. 1801–1863. Elsevier.

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.

Chen, X. (2013). STEM Attrition: College Students' Paths into and out of STEM Fields. Statistical Analysis Report. NCES 2014-001. *National Center for Education Statistics*.

Coviello, D., A. Ichino, and N. Persico (2014). Time allocation and task juggling. *American Economic Review 104*(2), 609–23.

Coviello, D., A. Ichino, and N. Persico (2015). The inefficiency of worker time use. *Journal of the European Economic Association 13*(5), 906–947.

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.

Epple, D. and R. E. Romano (2011). Chapter 20 - Peer Effects in Education: A Survey of the Theory and Evidence. Volume 1 of *Handbook of Social Economics*, pp. 1053–1163. North-Holland.

Griffith, A. L. (2010). Persistence of women and minorities in STEM field majors: Is it the school that matters? *Economics of Education Review 29*(6), 911–922.

Heckman, J. J., J. Stixrud, and S. Urzua (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics 24*(3), 411–482.

Imbens, G. W. and D. B. Rubin (2015). Causal inference in statistics, social, and biomedical sciences. *Cambridge University Press*.

Isphording, I., P. Qendrai, et al. (2019). Gender Differences in Student Dropout in STEM. *IZA Research Report87*.

Kahn, S. and D. Ginther (2018). Women and Science, Technology, Engineering, and Mathematics (STEM). In *The oxford handbook of women and the economy.*

Krashinsky, H. (2014). How would one extra year of high school affect academic performance in university? Evidence from an educational policy change. *Canadian Journal of Economics/Revue canadienne d'économique 47*(1), 70–97.

MacKinnon, J. G. and M. D. Webb (2017). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics 32*(2), 233–254.

Marcus, J. and V. Zambre (2018). The effect of increasing education efficiency on university enrollment: Evidence from administrative data and an unusual schooling reform in Germany. *Journal of Human Resources*, 1016–8324R.

OECD (2018). Education at a Glance 2018: OECD Indicators. *OECD Publishing, Paris*.

Peri, G., K. Shih, and C. Sparber (2015). STEM workers, H-1B visas, and productivity in US cities. *Journal of Labor Economics 33*(S1), S225–S255.

Pischke, J.-S. (2007). The impact of length of the school year on student performance and earnings: Evidence from the German short school years. *The Economic Journal 117*(523), 1216–1242.

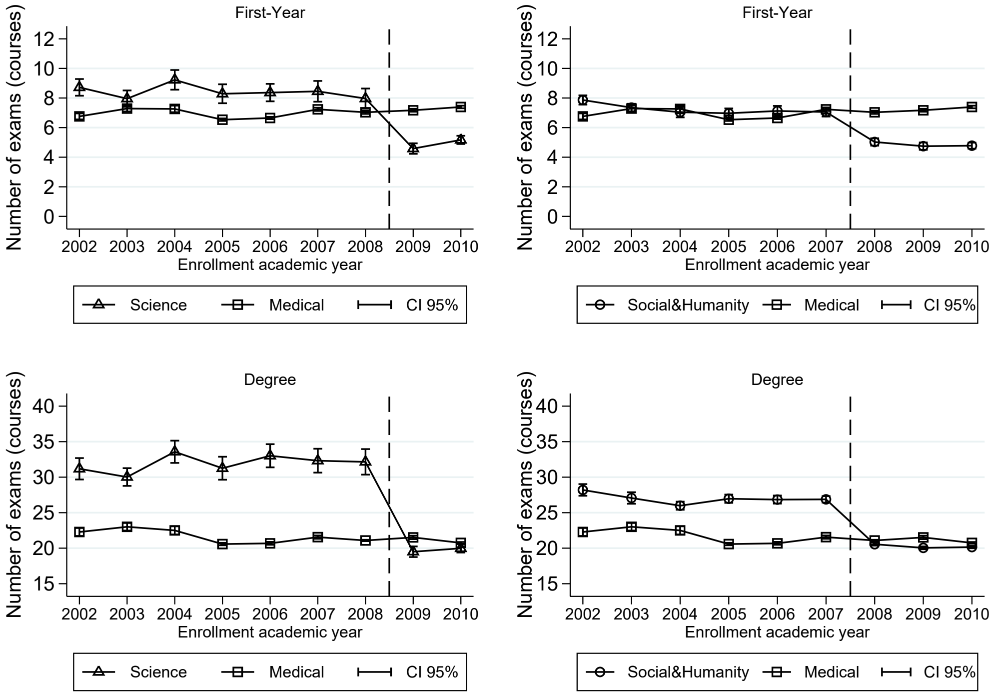
Stinebrickner, R. and T. Stinebrickner (2013). A major in science? Initial beliefs and final outcomes for college major and dropout. *The Review of Economic Studies 81*, 426–472.

Triventi, M. and P. Trivellato (2009). Participation, performance and inequality in Italian higher education in the 20th century. *Higher Education 57*(6), 681–702.

Tuckman, B. W. (1998). Using tests as an incentive to motivate procrastinators to study. *The Journal of Experimental Education 66*(2), 141–147.

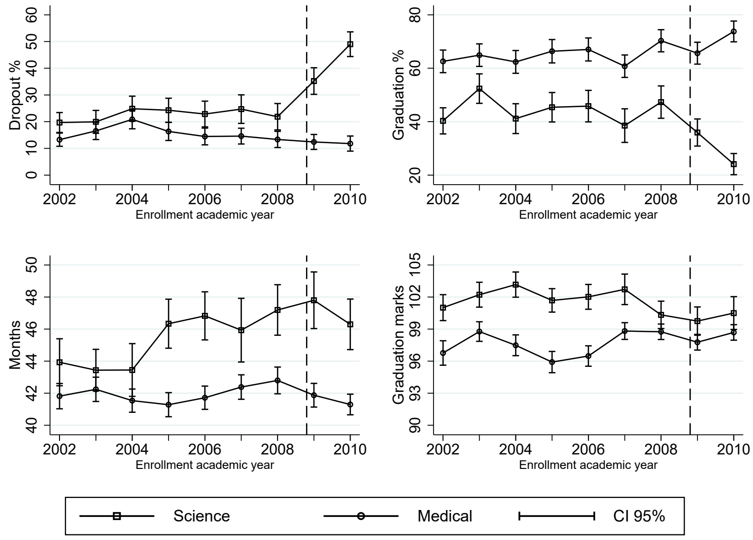
Figures

Fig. 1. **Number of exams (courses)**



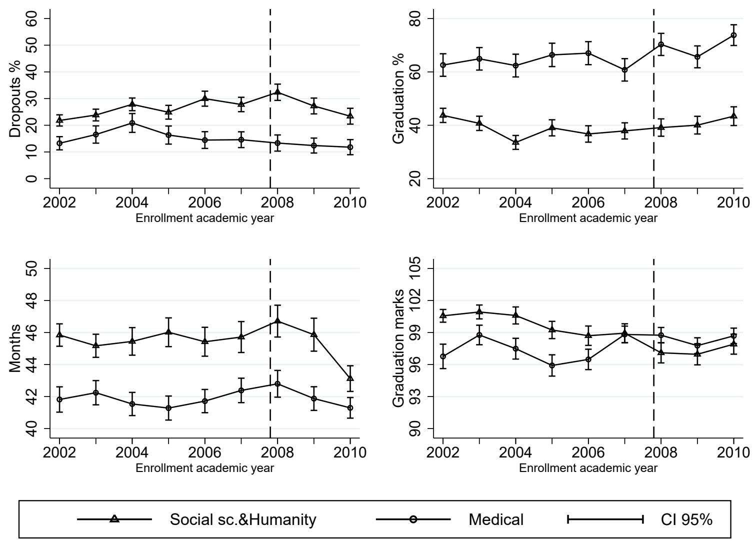
Notes: Fig. 1 plots the average number of exams (courses) taken by students who completed their degrees within 45 months, across scientific fields, over the academic years (at enrollment). The figures on top highlight the number of exams in the first year, while the figures in the bottom show the number of exams for the entire duration of the degree program. The sample of science and medical degrees comprises 2,885 graduates (605 in science), while the sample of social science, humanity, and medical degrees comprises 3,724 graduates (1,444 in social science and humanity). Vertical lines represent the introduction of the reform for programs in social sciences and humanities (in 2008–09) and science (in 2009–10). Confidence intervals are at the 95% level.

Fig. 2. **Descriptive statistics of academic outcomes (Science vs. Medical)**



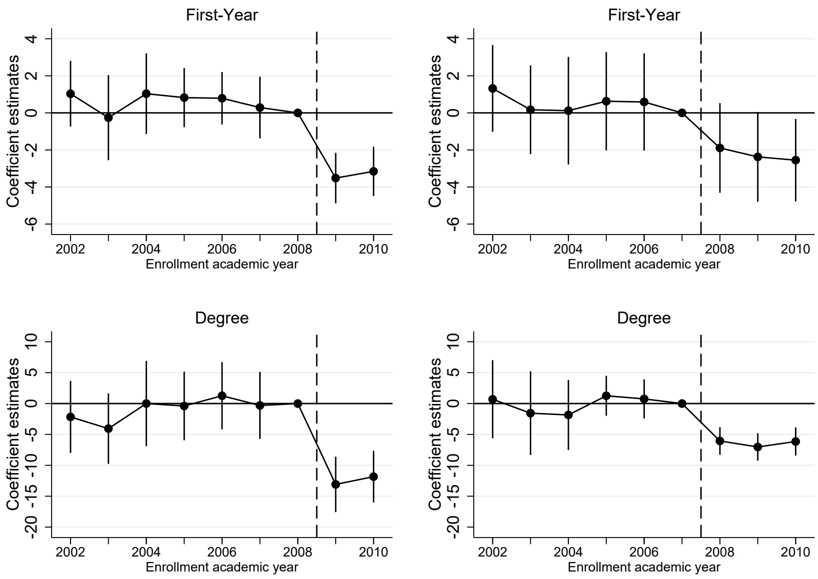
Notes: Fig. 2 highlights the averages of outcomes of students in science degree (square) and medical school (circle). The figure in the upper-left presents the results for first-year drop-out rates, N: 7,819. The figure in the upper-right presents the results for graduation rates, N: 7,277. The figure in the bottom-left presents the results for the duration (in months) of degree completion of graduates, N: 4,059. The figure in the bottom-right presents the results for the final marks of graduates, N: 4,059. Confidence intervals are at the 95% level.

Fig. 3. **Descriptive statistics of academic outcomes (Social sc. Humanities vs. Medical)**



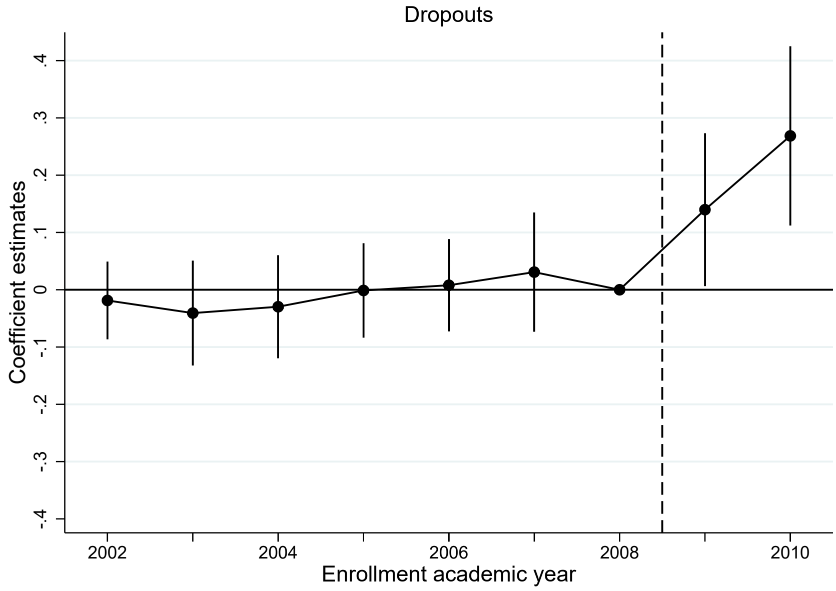
Notes: Fig. 3 highlights the averages of outcomes of students in social sciences and humanities degrees (triangle) and medical school (circle). The figure in the upper-left presents the results for first-year drop-out rates, N: 12,383. The figure in the upper-right presents the results for graduation rates, N: 11,523. The figure in the bottom-left presents the results for the duration (in months) of degree completion of graduates, N: 5877. The figure in the bottom-right presents the results for the final marks of graduates, N: 5,877. Confidence intervals are at the 95% level.

Fig. 4. **Event study specification: number of exams (courses)**



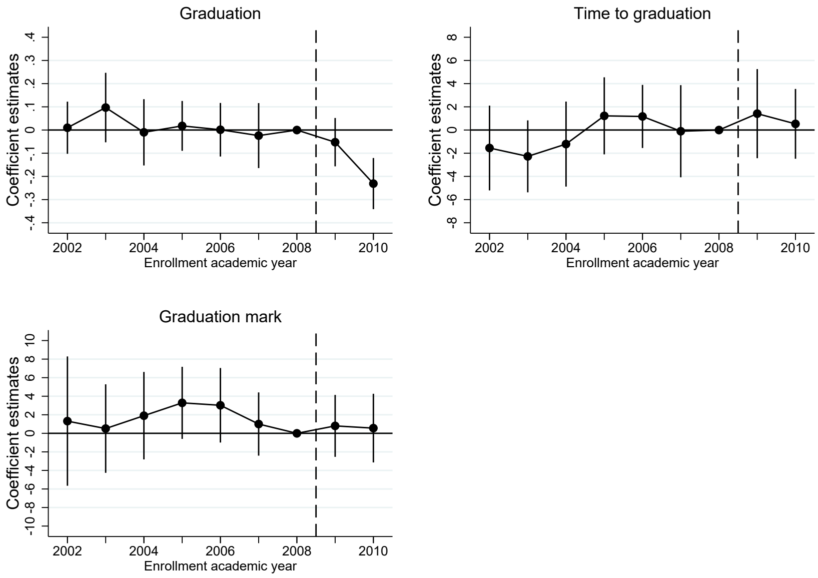
Notes: Fig. 4 highlights the coefficient estimates of δ_k specified in Equation (2) for the sample of graduates. The dependent variable is the number of exams that students passed. The figures at the top highlight the results for the first-year students, and the figures at the bottom for the entire period of degrees. The sample of science and medical degrees comprises 2,885 graduates (605 in science), while the sample of social science, humanity, and medical degrees comprises 3,724 graduates (1,444 in social science and humanity). Standard errors are clustered at the degree course and academic year level. Confidence intervals are at the 90% level.

Fig. 5. **Event study specification: Science vs. Medical**



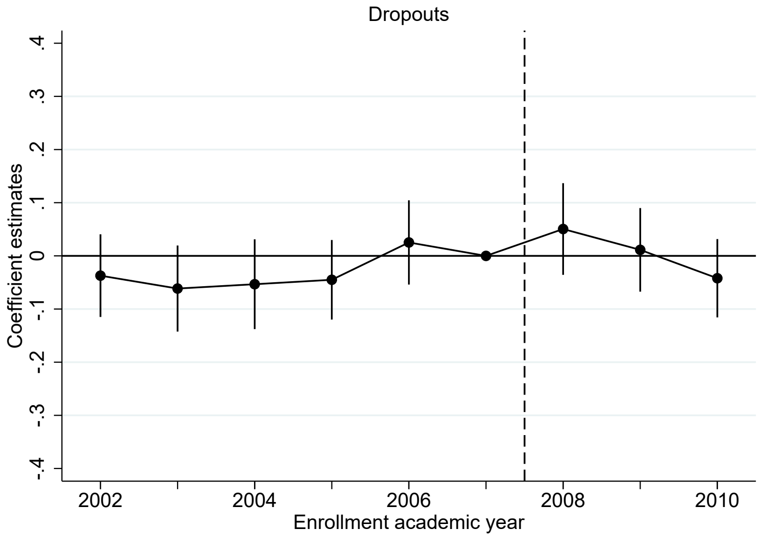
Notes: Fig. 5 highlights the coefficient estimates of δ_k specified in Equation (2) for the sample of degrees in science (treatment) and medical school (control), N: 7,819. The dependent variable is the drop-out probability at the end of the first year. The control variables included are age, gender, type of high school, final high school marks, and the region of high school. Standard errors are clustered at the degree course and academic year level. Confidence intervals are at the 90% level.

Fig. 6. **Event study specification: Science vs. Medical**



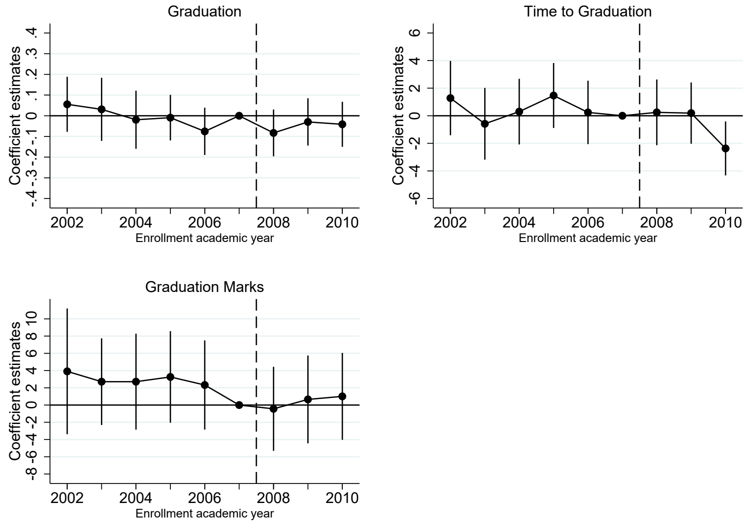
Notes: Fig. 6 highlights the coefficient estimates of δ_k specified in Equation (2) for the sample of degrees in science (treatment) and medical school (control). The figure in the upper-left shows the results for the probability of graduating from degrees, N: 7,277. The figure in the upper-right shows the results for the time to complete the degree (measured in months starting from enrollment), while the figure in the bottom-left shows the results for the final graduation marks of graduates. The estimation samples comprise only graduates in the figures in the upper-right and bottom-left, N: 4,059. The control variables included are age, gender, type of high school, final high school marks, and the region of high school. Standard errors are clustered at the degree course and academic year level. Confidence intervals are at the 90% level.

Fig. 7. **Event study specification: Social Sciences and Humanities vs. Medical**



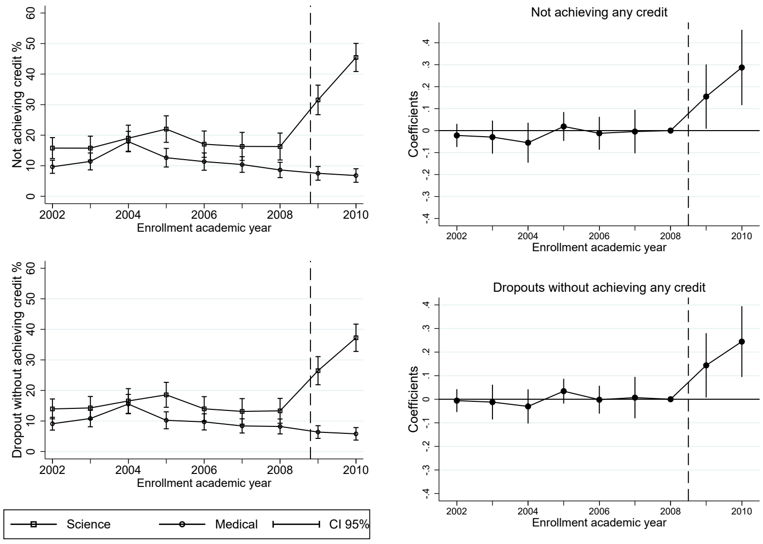
Notes: Fig. 7 highlights the coefficient estimates of δ_k specified in Equation (2) for the sample of degrees in social sciences, humanities and medical school, N: 12,383.The dependent variable is the probability of dropping out at the end of the first year. The control variables included are age, gender, type of high school, final high school marks, and the region of high school. Standard errors are clustered at the degree course and academic year level. Confidence intervals are at the 90% level.

Fig. 8. **Event study specification: Social Sciences and Humanities vs. Medical**



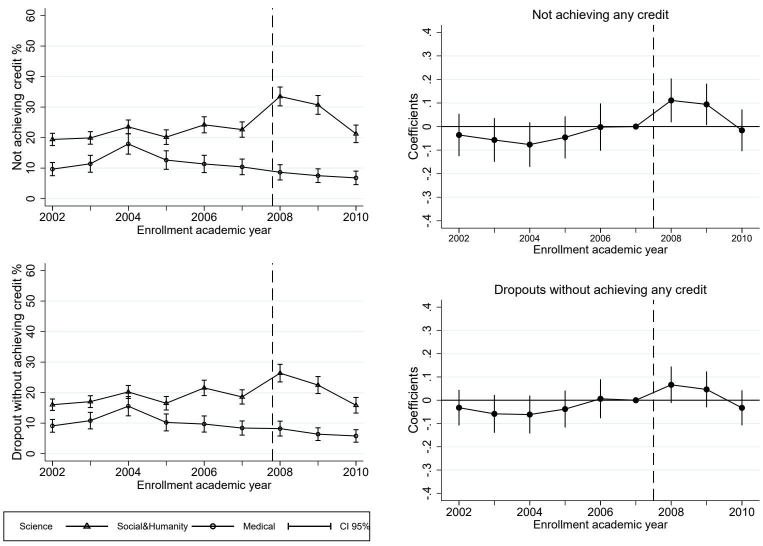
Notes: Fig. 8 highlights the coefficient estimates of δ_k specified in Equation (2) for the sample of degrees in social sciences, humanities and medical school. T. The figure in the upper-left shows the results for the probability of graduating from degrees, N: 11,523. The figure in the upper-right shows the results for the time to complete the degree (measured in months starting from enrollment), while the figure in the bottom-left shows the results for the final graduation marks of graduates. The estimation samples comprise only graduates in the figures in the upper-right and bottom-left, N: 5,877. The control variables include are age, gender, type of high school, final high school marks, and the region of high school. Standard errors are clustered at the degree course and academic year level. Confidence intervals are at the 90% level.

Fig. 9. **Event study specification: Science vs. Medical**



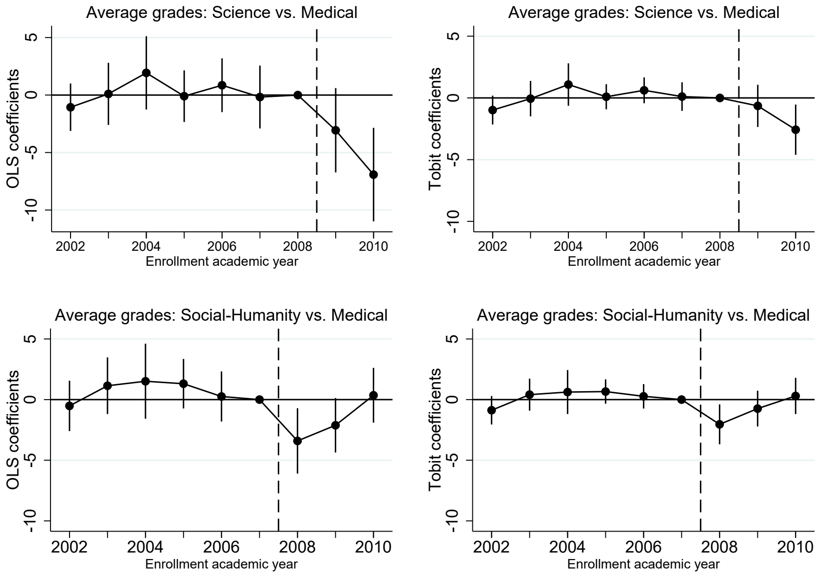
Notes: Fig. 9 reports the findings on the probability of not achieving any credit (figures on top) and the probability of dropping out without achieving credits (figures on bottom). The figures on the right highlight the coefficient estimates of δ_k specified in Equation (2) for the sample of degrees in science and medical school. N: 7,819. The control variables included are age, gender, type of high school, final high school marks, and the region of high school. Standard errors are clustered at the degree course and academic year level. Confidence intervals are at the 90% level.

Fig. 10. **Event study specification: Social Sciences and Humanities vs. Medical**



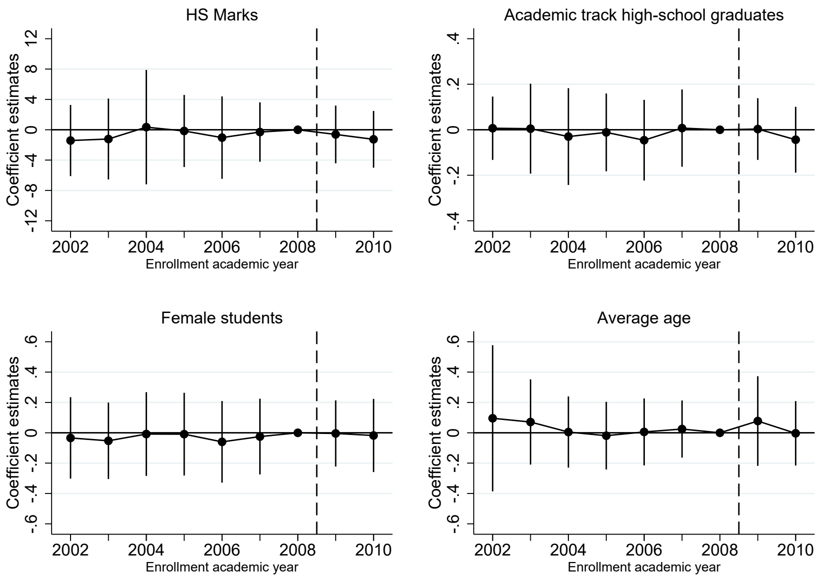
Notes: Fig. 10 reports the findings on the probability of not achieving any credit (figures on top) and the probability of dropping out without achieving credits (figures on bottom). The figures on the right highlight the coefficient estimates of δ_k specified in Equation (2). N: 12,383. The control variables included are age, gender, type of high school, final high school marks, and the region of high school. Standard errors are clustered at the degree course and academic year level. Confidence intervals are at the 90% level.

Fig. 11. **Event study specification on grades**



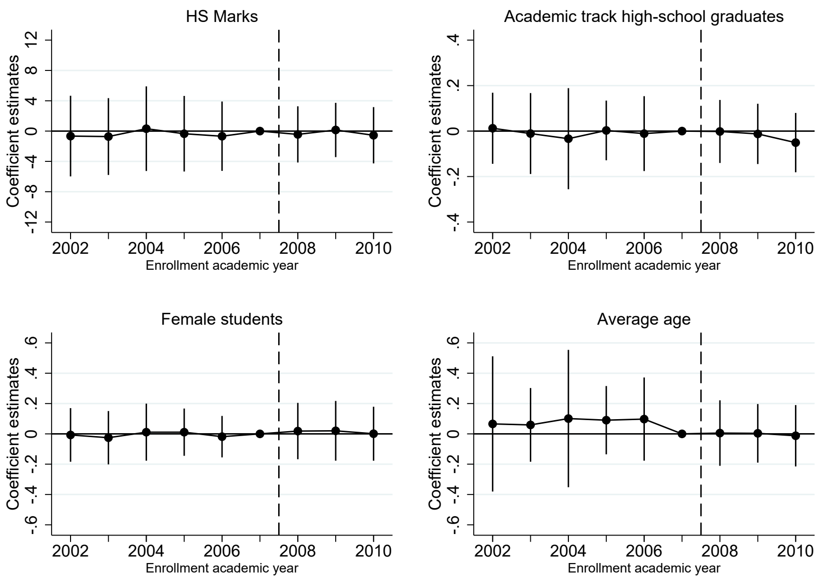
Notes: The figures on left highlight the coefficient estimates of δ_k specified in Equation (2) for both of the estimation samples. The figures on right highlight the coefficient estimates obtained from Tobit regressions outlined in footnote 27. The sample of degrees in science and medical school has 7,819 observations, while the sample of degrees in social sciences, humanities and medical school contains 12,383. The outcome variable is the grades of students (conditional on passing exam). The control variables included are age, gender, type of high school, final high school marks, and the region of high school. Standard errors are clustered at the degree course and academic year level. Confidence intervals are at the 90% level.

Fig. 12. **Event study specification for students' composition: Science vs. Medical**



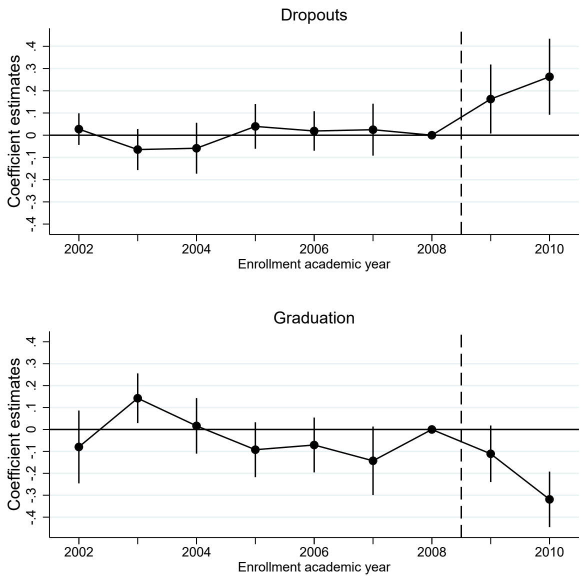
Notes: Fig. 12 highlights the coefficient estimates of δ_k specified in Equation (2) for the matched sample of degrees in science and medical school, N: 4,420. Standard errors are clustered at the degree course and academic year level. Confidence intervals are at the 90% level.

Fig. 13. **Event study specification for students' composition: Social sciences and Humanities vs. Medical**



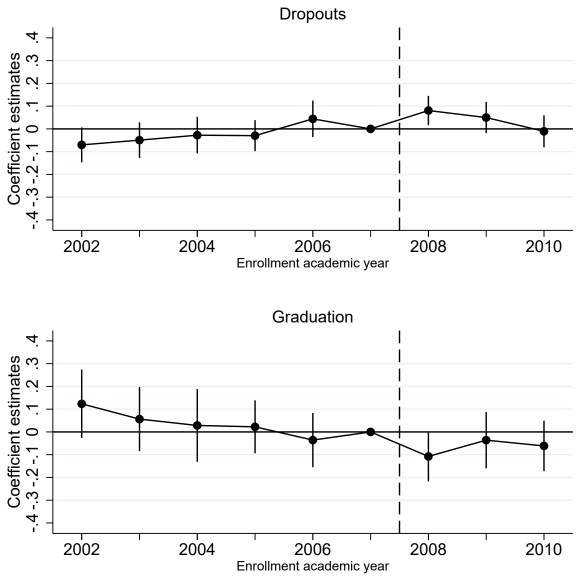
Notes: Fig. 13 highlights the coefficient estimates of δ_k specified in Equation (2) for the matched sample of degrees in social sciences, humanities and medical school, N: 9,787.Standard errors are clustered at the degree course and academic year level. Confidence intervals are at the 90% level.

Fig. 14. **Event-study specification (matched sample): Science vs. Medical**



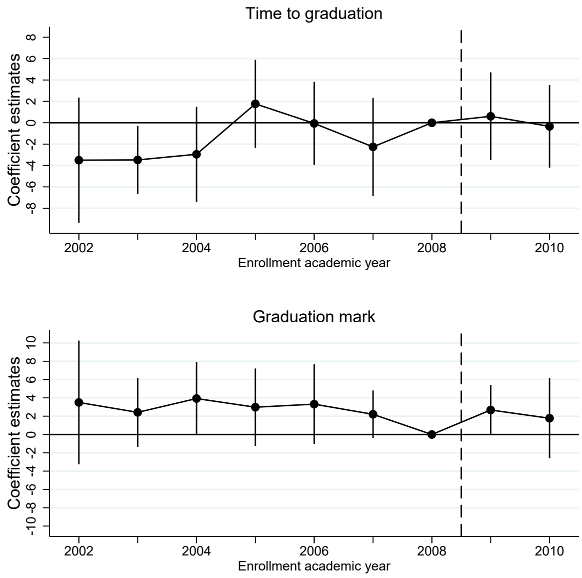
Notes: Fig. 14 highlights the coefficient estimates of δ_k specified in Equation (2) for the matched sample of degrees in science and medical school. The dependent variable in the upper figure is the first-year drop-out rate, N: 4,420. The dependent variable in the lower figure is the graduation rate, N: 4,224. The difference in the number of observations across figures is due to the cap that we introduce to time to graduation for older cohorts. The control variables included are age, gender, type of high school, final high school marks, and the region of high school. Standard errors are clustered at the degree course and academic year level. Confidence intervals are at the 90% level.

Fig. 15. **Event study specification (matched sample): Social Sciences and Humanities vs. Medical**



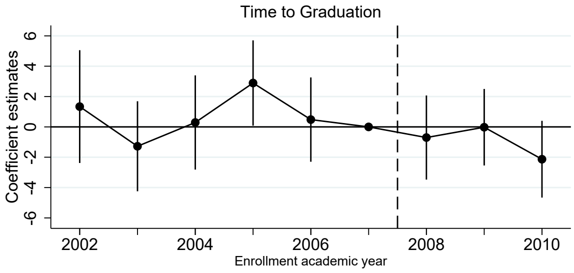
Notes: Fig. 15 highlights the coefficient estimates of δ_k specified in Equation (2) for the matched sample of degrees in social sciences, humanities and medical school, N: 9,787. The control variables included are age, gender, type of high school, final high school marks, and the region of high school. Standard errors are clustered at the degree course and academic year level. Confidence intervals are at the 90% level.

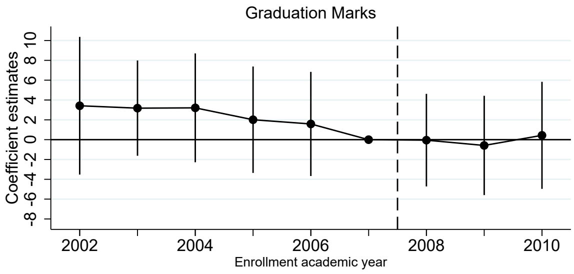
Fig. 16. **Event study specification (matched sample): Science vs. Medical**



Notes: Fig. 16 highlights the coefficient estimates of δ_k specified in Equation (2) for the matched sample of graduates of degrees in science and medical school, N: 1,729. The control variables included are age, gender, type of high school, final high school marks, and the region of high school. Standard errors are clustered at the degree course and academic year level. Confidence intervals are at the 90% level.

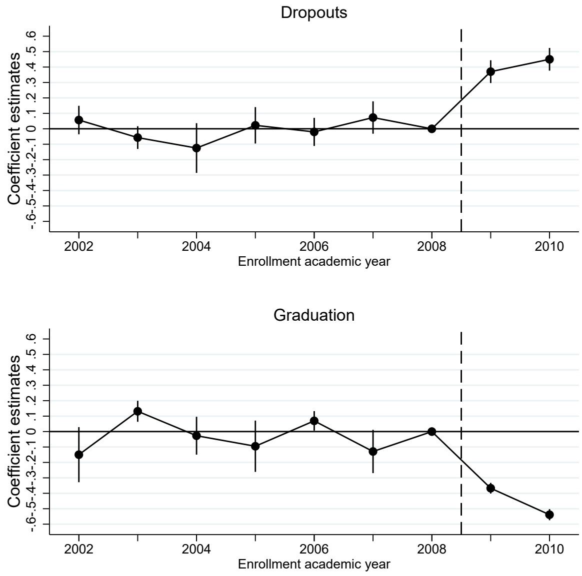
Fig. 17. **Event study specification (matched sample): Social Sciences and Humanities vs. Medical**





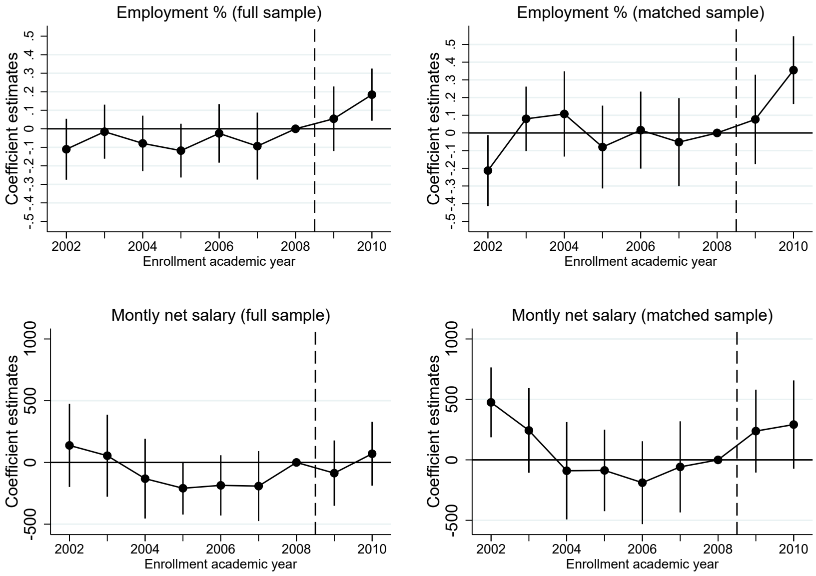
Notes: Fig. 17 highlights the coefficient estimates of δ_k specified in Equation (2) for the matched sample of graduates of degrees in social sciences, humanities and medical school, N: 2,797. The control variables included are age, gender, type of high school, final high school marks, and the region of high school. Standard errors are clustered at the degree course and academic year level. Confidence intervals are at the 90% level.

Fig. 18. **Event study specification (matched sample): Biotechnology vs. Medical**



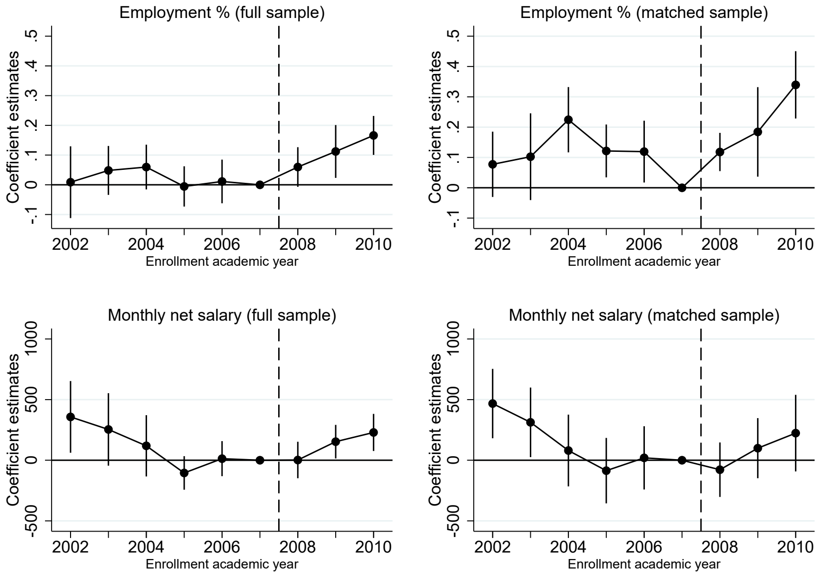
Notes: Fig. 18 highlights the coefficient estimates of δ_k specified in Equation (2) for the matched sample of students in biotechnology and medical school, N: 1,208. The control variables included are age, gender, type of high school, final high school marks, and the region of high school. Standard errors are clustered at the degree course and academic year level. Confidence intervals are at the 90% level.

Fig. 19. **Event study specification : Science vs. Medical**



Notes: Fig. 19 highlights the coefficient estimates of δ_k specified in Equation (2) for both full (left column) and matched (right column) samples of science and medical school. Dependent variables are the employment status (figures on top) and net monthly income (figures on bottom) at the moment of the interview. The number of observations in the upper-left figure is 2,898 (636 treated), in the upper-right figure it is 1,044 (311 treated), in the bottom-left figure it is 2,370, and in the bottom-right figure it is 830. The reason for having a lower number of observations in the figures on monthly salaries is because those regressions are conditional on employment. Standard errors are clustered at the degree course and academic year level. Confidence intervals are at the 90% level.

Fig. 20. **Event study specification: Social Sciences-Humanities vs. Medical**



Notes: Fig. 20 highlights the coefficient estimates of δ_k specified in Equation (2) for both the full (left column) and matched (right column) samples of social science, humanity, and medical school. Dependent variables are the employment status (figures on top) and net monthly income (figure on bottom) at the moment of the interview. The number of observations in the upper-left figure is 3,941 (1,751 treated), in the upper-right figure it is 2,814 (869 treated), in the bottom-left figure it is 3,181, and in the bottom-right figure it is 2,254. The reason for having fewer number of observations in the figures on monthly salaries is because those regressions are conditional on employment. Standard errors are clustered at the degree course and academic year level. Confidence intervals are at the 90% level.

Tables

Table 1. **Mean statistics of the observable characteristics of students by scientific degrees**

	First estimation sample				Second estimation sample			
	2002/03–2008/09		2009/10–2010/11		2002/03–2007/08		2008/09–2010/11	
	Medical	Science	Medical	Science	Medical	Socio-Human	Medical	Socio-Human
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female %	69.5	43.3	64.3	54.4	70.0	60.7	65.0	59.6
	(.46)	(.49)	(.48)	(.50)	(.46)	(.48)	(.47)	(.49)
Age	22.6	21.1	22.3	20.3	22.6	23.4	22.4	22.4
	(5.80)	(4.83)	(5.77)	(4.3)	(5.72)	(7.4)	(5.9)	(7.1)
HS. Marks	74.2	78.3	74.8	74.7	74.10	77.0	74.8	75.2
	(11.4)	(12.5)	(10.8)	(11.1)	(11.4)	(12.5)	(11.0)	(11.4)
HS-Science %	19.5	24.5	32.1	28.4	18.3	17.5	30.4	19.1
	(.39)	(.43)	(.46)	(.45)	(.39)	(.38)	(.46)	(.39)
HS-Technical %	32.5	41.4	32.4	32.8	32.7	36.3	32.0	37.4
	(.46)	(.49)	(.46)	(.47)	(.47)	(.48)	(.46)	(.48)
HS-Classical %	6.0	4.6	6.2	8.0	6.0	7.5	6.0	8.2
	(.23)	(.21)	(.24)	(.27)	(.24)	(.26)	(.24)	(.27)
HS-Linguistic %	10.4	7.3	8.7	9.1	10.2	11.4	9.5	(13.1)
	(.30)	(.26)	(.28)	(.28)	(.30)	(.32)	(.30)	(.33)
HS-Professional %	16.3	7.6	7.2	6.0	16.7	10.1	9.3	7.3
	(.37)	(.26)	(.26)	(.23)	(37)	(.30)	(.29)	(.26)
HS-Others %	15.2	14.4	13.0	15.5	15.8	17.0	12.5	14.6
	(.36)	(.35)	(.33)	(.36)	(.36)	(.37)	(.33)	(.35)
Piemonte %	72.4	91.0	71.3	88.0	73.5	84.7	69.1	81.2
	(.44)	(.28)	(.45)	(.32)	(.44)	(.36)	(.46)	(.39)
N	3,718	2,263	1,032	806	3,231	8,148	1,519	2,807

Notes: The table reports the mean statistics of the observable characteristics of students by scientific degrees. Columns (1), (2), (3), and (4) show the statistics for the first estimation sample that consists of science and medical school degrees. The columns (5), (6), (7), and (8) report the statistics for the second estimation sample that consists of social sciences, humanities (Socio-Human), and medical school degrees. Standard deviations are in parenthesis.

Table 2. Main results (Science vs. Medical)

	Drop		Grad.		Months		Marks	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Treated * Post_{2t}$	0.284***	0.255***	-0.282***	-0.228***	0.457	0.026	-2.268	-1.109
	(0.084)	(0.088)	(0.059)	(0.060)	(1.543)	(1.467)	(2.687)	(2.153)
$Treated * Post_{1t}$	0.140*	0.129*	-0.082	-0.051	1.380	0.755	-2.103	-0.948
	(0.072)	(0.073)	(0.061)	(0.056)	(2.161)	(1.989)	(2.491)	(1.950)
Treated	0.088***	0.106***	-0.215***	-0.269***	4.544***	5.614***	4.082***	1.540
	(0.021)	(0.022)	(0.034)	(0.031)	(0.784)	(0.756)	(1.262)	(1.077)
$Post_{2t}$	-0.029**	-0.009	0.078***	0.047**	-0.771	-0.609	1.154	0.645
	(0.014)	(0.016)	(0.027)	(0.021)	(0.695)	(0.503)	(2.299)	(1.666)
$Post_{1t}$	-0.023	-0.017	-0.003	-0.019	-0.192	-0.074	0.239	-0.167
	(0.017)	(0.015)	(0.029)	(0.020)	(0.797)	(0.561)	(2.254)	(1.634)
Female		0.004		0.046***		-0.382		-0.740**
		(0.014)		(0.017)		(0.251)		(0.360)
Age		0.003*		-0.005***		-0.101**		0.156***
		(0.001)		(0.001)		(0.041)		(0.028)
HS. Marks		-0.003***		0.007***		-0.182***		0.289***
		(0.001)		(0.001)		(0.013)		(0.016)
Professional		0.130***		-0.200***		1.534***		-4.461***
		(0.027)		(0.027)		(0.433)		(0.510)
Technical		0.069***		-0.087***		1.446***		-2.982***
		(0.019)		(0.023)		(0.319)		(0.271)
Classical		0.058**		-0.122***		0.608		-0.911**
		(0.024)		(0.031)		(0.543)		(0.439)
Linguistic		0.070**		-0.081***		2.045***		-3.460***
		(0.027)		(0.029)		(0.511)		(0.444)
Other HS.		0.042**		-0.084***		1.256**		-2.870***
		(0.018)		(0.026)		(0.501)		(0.635)
Region fixed-effect	No	Yes	No	Yes	No	Yes	No	Yes
Observations	4,980	4,980	4,781	4,781	2,667	2,667	2,667	2,667

Notes: Table 2 reports the estimation results from Equation 1 for the sample of degrees in science (*Treated*) and medical school (control). The sample covers the academic years at enrollment from 2005–06 to 2010–11. Columns (1), (3), (5), (7) show the results when the control variables are not included into regressions, while columns (2), (4), (6), (8) show the results when the control variables are included. (1)-(2) are the probability of dropping out in the first year. (3)-(4) are the probability of graduating. (5)-(6) are the duration (in months) of degree completion of graduates. (7)-(8) are the final graduation marks of graduates. HS. Marks stand for the final score of the high school exam. $Post_{2t}$ is a dummy that takes the value of 1 if the academic year is 2010–11 at enrollment, $Post_{1t}$ is a dummy that takes the value of 1 if the academic year is 2009–10 at enrollment. The omitted category of high school types is scientific schools. Region fixed-effects stand for the region of the high school students that graduated from. Standard errors are in parentheses and clustered at the degree program and academic year level. * $p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table 3. Main results (Social sc. and Humanities vs. Medical)

	Drop		Grad.		Months		Marks	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Treated * Post_{2t}$	0.024	0.015	-0.016	0.002	0.270	-0.447	-2.274	-1.241
	(0.031)	(0.027)	(0.052)	(0.039)	(1.187)	(0.911)	(3.495)	(2.872)
$Treated * Post_{1t}$	0.066*	0.052	-0.071	-0.051	0.182	-0.362	-3.153	-2.278
	(0.034)	(0.035)	(0.051)	(0.037)	(1.333)	(1.039)	(3.217)	(2.747)
Treated	0.124***	0.130***	-0.240***	-0.276***	3.720***	5.185***	1.462	-0.883
	(0.020)	(0.020)	(0.030)	(0.027)	(0.680)	(0.621)	(1.773)	(1.524)
Pos_{2t}	-0.027	-0.017	0.011	-0.011	0.069	0.377	0.673	0.331
	(0.020)	(0.018)	(0.032)	(0.024)	(0.777)	(0.483)	(2.400)	(1.847)
$Post_{1t}$	-0.017	-0.006	0.058	0.040	0.992	1.074*	1.650	1.580
	(0.030)	(0.032)	(0.045)	(0.035)	(0.921)	(0.629)	(2.000)	(1.481)
Female		-0.024**		0.027**		0.289		0.566
		(0.011)		(0.012)		(0.246)		(0.461)
Age		0.005***		-0.011***		-0.032		0.295***
		(0.001)		(0.001)		(0.035)		(0.040)
HS. Marks		-0.005***		0.010***		-0.207***		0.322***
		(0.000)		(0.001)		(0.017)		(0.017)
Professional		0.154***		-0.255***		3.348***		-4.834***
		(0.026)		(0.034)		(0.732)		(0.453)
Technical		0.091***		-0.133***		2.146***		-3.649***
		(0.020)		(0.031)		(0.342)		(0.314)
Classical		0.068***		-0.081***		0.909		0.858
		(0.024)		(0.027)		(0.663)		(0.592)
Linguistic		0.085***		-0.135***		3.363***		-2.123***
		(0.028)		(0.039)		(0.623)		(0.530)
Other HS.		0.064***		-0.118***		2.597***		-2.407***
		(0.021)		(0.027)		(0.451)		(0.605)
Observations	6,834	6,834	6,497	6,497	3,231	3,231	3,231	3,231

Notes: Table 3 reports the estimation results from Equation 1 for the sample of degrees in social sciences and humanities ($Treated$) and medical school (control). The sample covers the academic years at enrollment from 2005–06 to 2009–10. Columns (1), (3), (5), (7) show the results when the control variables are not included into regressions, while columns (2), (4), (6), (8) show the results when the control variables are included. (1)-(2) are the probability of dropping out in the first year. (3)-(4) are the probability of graduating. (5)-(6) are the duration (in months) of degree completion of graduates. (7)-(8) are the final graduation marks of graduates. HS. Marks stand for the final score of the high school exam. $Post_{2t}$ is a dummy that takes the value of 1 if the academic year is 2009–10 at enrollment, $Post_{1t}$ is a dummy that takes the value of 1 if the academic year is 2008–09 at enrollment. The omitted category of high school types is scientific schools. Region fixed-effects stand for the region of the high school that students graduated from. Standard errors are in parentheses and clustered at the degree program and academic year level. * $p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table 4. **Summary statistics in matched samples before and after the reform**

	(1)	(2)	(3)	(4)	(5)	(6)
	Before reform (2005/06–2008/09)			After reform (2009/10–2010/11)		
	Treated	Control	Std. dif.	Treated	Control	Std. dif.
Panel-A: Science vs. Medical						
Academic track%	0.41	0.41	-0.00	0.50	0.53	-0.06
High-school marks	77.26	76.75	0.04	73.62	73.62	0.00
Female%	0.56	0.59	-0.06	0.57	0.58	-0.02
Age	19.29	19.24	0.06	19.42	19.33	0.04
Dropout%	0.27	0.16	0.27	0.41	0.12	0.70
Graduation%	0.46	0.69	-0.48	0.33	0.74	-0.82
Num. of observations	608	1,028		493	822	
	Before reform (2005/06–2007/08)			After reform (2008/09–2009/10)		
	Treated	Control	Std. dif.	Treated	Control	Std. dif.
Panel-B: Social Sc.-Humanity vs. Medical						
Academic track%	0.31	0.30	0.01	0.41	0.41	0.02
High-school marks	76.53	76.63	-0.01	74.60	74.23	0.04
Female%	0.68	0.71	-0.06	0.63	0.63	0.00
Age	19.65	19.47	0.10	19.50	19.40	0.07
Dropout%	0.25	0.15	0.26	0.27	0.13	0.38
Graduation%	0.50	0.69	-0.40	0.47	0.72	-0.52
Num. of observations	1,181	2,012		901	1,476	

Notes: Table 4 reports the summary statistics of variables in matched samples. Panel A shows the results for the sample of science and medical school students, while the Panel B shows them for the sample social science, humanities and medical school students. Columns (3) and (6) present the standardized mean differences calculated by dividing the mean difference between the outcomes of treated and control units by the standard deviation of the outcome.

Table 5. Heterogeneous effects on first-year dropout and graduation rates (Science vs. Medical)

Heterogeneous groups	Female		Academic		High-Ability		High-Income	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variables	Dropout	Grad.	Dropout	Grad.	Dropout	Grad.	Dropout	Grad.
$Treated * Post_2 * Het$	-0.047	0.095	-0.122*	0.006	-0.167**	0.215**	-0.111	0.132
	(0.083)	(0.083)	(0.067)	(0.082)	(0.083)	(0.098)	(0.119)	(0.127)
$Treated * Post_1 * Het$	-0.029	0.096	-0.072	0.031	-0.136	0.245**	-0.007	0.032
	(0.078)	(0.087)	(0.104)	(0.103)	(0.100)	(0.110)	(0.107)	(0.089)
$Treated * Post_2$	0.256**	-0.303***	0.288**	-0.252**	0.249**	-0.264***	0.268**	-0.285**
	(0.110)	(0.083)	(0.118)	(0.101)	(0.096)	(0.076)	(0.133)	(0.110)
$Treated * Post_1$	0.154	-0.082	0.167	-0.053	0.152*	-0.055	0.135	-0.042
	(0.102)	(0.072)	(0.108)	(0.083)	(0.088)	(0.079)	(0.091)	(0.075)
Treated	0.062	-0.226***	0.080*	-0.258***	0.080**	-0.236***	0.062*	-0.217***
	(0.053)	(0.057)	(0.042)	(0.050)	(0.031)	(0.044)	(0.035)	(0.041)
Het	-0.076**	0.107***	-0.103***	0.115***	-0.121***	0.148***	-0.074**	0.060
	(0.033)	(0.031)	(0.034)	(0.040)	(0.024)	(0.044)	(0.035)	(0.042)
Treated*Het	0.044	-0.012	0.027	0.050	0.039	0.013	0.072	-0.045
	(0.053)	(0.053)	(0.047)	(0.059)	(0.056)	(0.062)	(0.046)	(0.057)
Post1	-0.136***	0.052	-0.132***	0.067	-0.086***	0.003	-0.125***	0.032
	(0.037)	(0.037)	(0.030)	(0.052)	(0.024)	(0.032)	(0.027)	(0.034)
Post2	-0.084**	0.158***	-0.087***	0.134***	-0.043*	0.073**	-0.048	0.106***
	(0.035)	(0.043)	(0.028)	(0.035)	(0.022)	(0.028)	(0.036)	(0.035)
Post1*Het	0.113**	-0.090**	0.119***	-0.129	0.087**	-0.056	0.145***	-0.085
	(0.048)	(0.041)	(0.036)	(0.079)	(0.034)	(0.078)	(0.042)	(0.055)
Post2*Het	0.088***	-0.131**	0.110**	-0.109**	0.092*	-0.061	0.044	-0.090*
	(0.031)	(0.062)	(0.044)	(0.046)	(0.048)	(0.048)	(0.063)	(0.054)
Observations	2,950	2,950	2,950	2,950	2,950	2,950	2,946	2,946

Notes: Table 5 reports the estimation results from Equation 3 for the sample obtained from exact matching for the degrees in science (*Treated*) and medical school (control). The sample covers the academic years at enrollment from 2005–06 to 2010–11. *Het* stands for heterogeneous groups, which takes the value of 1 if the student is female in (1)-(2), if the student graduated from an academic track school in (3)-(4), if the student's final high school mark is in the top 25% in (5)-(6), if the tuition fee that the student is supposed to pay is in the top 25% in (7)-(8). $Post_{2t}$ is a dummy that takes the value of 1 if the academic year is 2010–11 at enrollment, $Post_{1t}$ is a dummy that takes the value of 1 if the academic year is 2009–10 at enrollment. All regressions include control variables on gender, age, type of high school, final high school marks, and the region of the high school that students graduated from. Standard errors are in parentheses and clustered at the degree program and academic year level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6. Heterogeneous effects on first-year dropout and graduation rates (Social sc. and Humanity vs. Medical)

Heterogeneous groups	Female		Academic		High-Ability		High-Income	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variables	Dropout	Grad.	Dropout	Grad.	Dropout	Grad.	Dropout	Grad.
$Treated * Post_2 * Het$	-0.054	-0.133*	-0.094*	0.111*	0.070	-0.027	-0.094	-0.023
	(0.063)	(0.076)	(0.053)	(0.064)	(0.085)	(0.087)	(0.064)	(0.078)
$Treated * Post_1 * Het$	-0.015	-0.046	-0.058	0.186***	0.016	-0.187**	0.062	-0.048
	(0.053)	(0.058)	(0.054)	(0.052)	(0.049)	(0.085)	(0.071)	(0.070)
$Treated * Post_2$	0.067	0.054	0.079**	-0.086*	0.014	0.000	0.061**	-0.016
	(0.054)	(0.064)	(0.033)	(0.050)	(0.038)	(0.060)	(0.030)	(0.063)
$Treated * Post_1$	0.081	-0.072	0.095**	-0.178***	0.063**	-0.059	0.042	-0.081
	(0.051)	(0.061)	(0.037)	(0.042)	(0.031)	(0.047)	(0.042)	(0.051)
Treated	0.092*	-0.212***	0.090***	-0.205***	0.107***	-0.236***	0.092***	-0.207***
	(0.049)	(0.051)	(0.030)	(0.039)	(0.028)	(0.040)	(0.028)	(0.038)
Het	-0.033	0.015	-0.089***	0.118***	-0.081***	0.168***	-0.029	-0.011
	(0.037)	(0.035)	(0.029)	(0.027)	(0.027)	(0.049)	(0.029)	(0.036)
Treated*Het	-0.011	0.035	-0.017	0.046	-0.095**	0.204***	-0.011	0.048
	(0.048)	(0.043)	(0.033)	(0.039)	(0.038)	(0.057)	(0.042)	(0.048)
Post1	-0.034	0.044	-0.022	0.129***	-0.005	0.066*	-0.023	0.095***
	(0.043)	(0.049)	(0.027)	(0.036)	(0.025)	(0.034)	(0.030)	(0.033)
Post2	-0.083*	0.006	-0.073**	0.060	-0.035	0.014	-0.077***	0.022
	(0.042)	(0.054)	(0.028)	(0.038)	(0.022)	(0.031)	(0.023)	(0.042)
Post1*Het	0.029	0.054	0.022	-0.132***	-0.045	0.061	0.028	-0.041
	(0.040)	(0.041)	(0.030)	(0.033)	(0.029)	(0.047)	(0.044)	(0.035)
Post2*Het	0.079*	0.007	0.086***	-0.100***	0.041	-0.102*	0.125***	-0.034
	(0.045)	(0.059)	(0.031)	(0.031)	(0.046)	(0.055)	(0.040)	(0.063)
Observations	5,567	5,567	5,567	5,567	5,567	5,567	5,556	5,556

Notes: Table 6 reports the estimation results from Equation 3 for the sample obtained from exact matching for the degrees in social sciences and humanities (*Treated*) and medical school (control). The sample covers the academic years at enrollment from 2005–06 to 2009–10. *Het* stands for heterogeneous groups, which takes the value of 1 if the student is female in (1)-(2), if the student graduated from an academic track school in (3)-(4), if the student's final high school mark is in the top 25% in (5)-(6), if the tuition fee that the student is supposed to pay is in the top 25% in (7)-(8). $Post_{2t}$ is a dummy that takes the value of 1 if the academic year is 2009–10 at enrollment, $Post_{1t}$ is a dummy that takes the value of 1 if the academic year is 2008–09 at enrollment. All regressions include control variables on gender, age, type of high school, final high school marks, and the region of the high school that students graduated from. Standard errors are in parentheses and clustered at the degree program and academic year level. * $p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

A. Università del Piemonte Orientale

The proposed analysis is on a small-/medium-sized public university located in the Nort-West of Italy, the Università del Piemonte Orientale.

Italy has 91 universities, of which 60 are public. All Italian universities have to comply with national laws that regulates their governing bodies – e.g. the teacher hiring procedures, the outline and organization of the degree programs – to make all academic qualifications issued comparable (for instance for the national competitions). Within this national regulatory framework, universities have strong autonomy in terms of teaching and research activity, and to some extent also financially. Italian universities are rather heterogeneous in size: the student population ranges from more than 100,000 at Rome La Sapienza (the largest university in Europe) to less than 10,000 in ten small universities mainly located in the Center and Southern areas of the country.

Although regulations have changed a lot since 2012, following the so-called "Gelmini Reform", in the years analyzed in the paper – 2002-2011 – all universities were organized in faculties (Facoltà) that were entitled to organize and manage both the teaching and research activities.

UPO was founded in 1998 in a peripheral area that lacked university sites. During the years around the reform (2007/08-2008/09-2009/10) the number of students enrolled was about 10,000. In the recent academic years, more than 14,000 students were enrolled, 4,000 for the first time (freshmen).

Like most public universities in Italy, UPO is a generalist institution as it offers all degrees (first cycle, second cycle, Ph.D.) in all scientific areas: science, health care, social and humanities. The only degree programs not offered in UPO are architecture and engineering, which in Italy are often offered by specialized institutions, named Politecnici.

Concerning students' characteristics, UPO freshmen are comparable with their peers in terms of age, as approximately 70% enroll for the first time at the "typical" age (18 or 19 years), i.e. immediately after having achieved a high school degree. To sum up, we argue that except for the distribution by scientific areas, UPO is representative of the average

Italian university, or at least of small-/medium-sized institutions, which represent 50% of all Italian universities.

B. Results from wildbootstrapping

Table 7. Main Results

	Drop	Grad.	Marks	Months
	(1)	(2)	(3)	(4)
Panel-A: Science vs. Medical				
$Treated * POST_{2t}$	0.284***	-0.282**	-2.268	0.457
t stat.	(2.981)	(-3.481)	(-1.714)	(0.239)
p-value	(0.010)	(0.012)	(0.148)	(0.886)
$Treated * POST_{1t}$	0.140	-0.082	-2.103**	1.380
t stat.	(1.811)	(-1.240)	(-1.982)	(1.154)
p-value	(0.166)	(0.334)	(0.043)	(0.274)
N	4,980	4,781	2,667	2,667
Panel-B: Social sciences and Humanities vs. Medical				
$Treated * POST_{2t}$	0.024	-0.016	-2.274	0.270
t stat.	(0.922)	(-0.451)	(-2.781)	(0.475)
p-value	(0.302)	(0.705)	(0.269)	(0.612)
$Treated * POST_{1t}$	0.066*	-0.071	-3.153	0.182
t stat.	(2.096)	(-1.439)	(-2.919)	(0.186)
p-value	(0.053)	(0.205)	(0.224)	(0.853)
N	6,834	6,497	3,231	3,231

Notes: Table 7 reports the estimation results from Equation 1. Panel A shows the results for the estimation sample comprising science (treatment) and medical school (control) degrees, while Panel B shows the results for the sample of social science and humanity degrees (treatment) and medical school (control). (1) is the probability of dropping out in the first year. (2) is the probability of graduating. (3) is the final marks of graduates. (4) is the duration (in months) of degree completion of graduates. Regressions do not include control variables. P-values are computed by clustering standard errors at the degree program level and performing wild-bootstrapping. * $p < 0.10$, **$p < 0.05$, ***$p < 0.01$.