European Economic Review 146 (2022) 104164

Contents lists available at ScienceDirect

European Economic Review

journal homepage: www.elsevier.com/locate/eer

# When need meets merit: The effect of increasing merit requirements in need-based student aid

Veronica Minaya [b,*], Tommaso Agasisti [a], Massimiliano Bratti [c]

[a] Politecnico di Milano, Politecnico di Milano School of Management, Via Lambruschini 4b, 20156, Milan, Italy
[b] Community College Research Center, Teachers College, Columbia University, 525 W. 120th St. Box 174, New York, NY, 10027, United States
[c] Università degli Studi di Milano, and GLO, IZA and LdA, Department of Economics, Management and Quantitative Methods (DEMM), Università degli Studi di Milano, via Conservatorio 7, 20122, Milan (Italy)

ARTICLE INFO

*Keywords:*
Student financial aid
Merit-based requirements
University
Difference-in-differences
Italy

ABSTRACT

Merit requirements in need-based student aid may exacerbate inequality in higher education but at the same time improve efficiency of aid expenditure by increasing on-time graduation, for instance. Disentangling the effect of the two building blocks of student aid ("need" and "merit") is therefore of key interest to policy makers. In this paper, we seek to estimate the causal effect of tightening the academic requirements embodied in need-based student aid on short-term and long-term student academic performance. This is done leveraging a reform in an Italian region that increased by 40% (i.e. from 25 to 35 out of a maximum of 60) the number of credits to be earned in the first academic year to maintain aid eligibility. Using administrative data from an Italian public university mainly offering STEM degrees, this study reveals that tightening merit requirements had a statistically significant, positive effect on various dimensions of performance of the "average" aid recipient. However, an analysis of treatment heterogeneity unveils winners and losers from the policy: the positive effects are indeed concentrated among higher and medium-ability students, while lower-ability students receiving financial assistance are discouraged from continuing in their studies.

## 1. Introduction

Need-based financial aid is a widely used tool to help make college more affordable for low-income students. While need-based aid programs are initially based purely on financial need, most of these programs attach merit-based requirements for renewal. Merit-based requirements may increase the efficiency of aid expenditure but at the cost of reducing equality of educational opportunities (*equity-efficiency trade-off*). These requirements may increase the risk of losing funding because of failure to renew and consequently increase dropout rates. Therefore, it is important for policymakers to better understand the role of merit-based requirements in need-based aid and to address both efficiency and equity concerns.

Despite its importance, merit-based requirements for need-based aid programs have received little attention from researchers and policymakers. Need-based programs are designed as a "bundled" package, and in studies assessing the effect of financial aid on student performance, it is generally difficult to disentangle the effect of the "need" component from that of the "merit" component. Although the literature finds that minimum merit requirements in need-based aid may hurt the persistence of low-income students, arguing that

* Corresponding author.
*E-mail addresses:* vmm2122@tc.columbia.edu (V. Minaya), tommaso.agasisti@polimi.it (T. Agasisti), massimiliano.bratti@unimi.it (M. Bratti).

https://doi.org/10.1016/j.euroecorev.2022.104164
Received 15 June 2021; Received in revised form 12 May 2022; Accepted 16 May 2022
Available online 6 June 2022

merit requirements promote inequality in HE (Scott-Clayton and Schudde, 2020), there is little evidence on the effect of *increasing merit requirements.* An important exception is Montalbán (2019), who studied the causal effect of receiving the same grant amount under different intensities of merit requirements on student academic outcomes at a HE institution in Spain. He finds positive effects of being eligible for a grant on student performance when combined with stronger merit requirements. His results suggest that merit requirements attached to need-based aid may be an effective tool to improve student performance and aid effectiveness. We contribute to this scant literature by leveraging a reform introduced in 2011, when the regional government of the most populous Italian region (Lombardy, North-West Italy) raised the merit requirements for students to maintain their need-based grant. We draw from this reform to explore the consequences of tightening the merit requirements on need-based student aid for cohorts of students enrolled at the *Politecnico di Milano* (PoliMi, hereafter) after the reform.

In this paper, we estimate the causal impact on a set of short and long-run academic outcomes, such as the performance in the first year of bachelor studies (number of credits earned and average grades obtained in the first year, i.e. GPA), student retention and final graduation mark, probability of on-time graduation, i.e. within the normative time to completion (three years), or with at most one-year delay. Our analysis focuses on first-time-in-college students enrolled in a bachelor's degree (BA) at PoliMi between 2008 and 2013. We follow these cohorts of students until 2017.

We use a difference-in-differences (DID) approach to evaluate whether aid-eligible students in cohorts enrolled after the reform experienced changes in academic performance due to the tightening of the merit component of the need-based program (holding the financial need requirement constant). The reform may have changed the composition of students who apply for financial aid and encouraged cream-skimming. We assess and address the selection or "cream-skimming" problem and evaluate sensitivity to alternative specifications.

The main results of this study lead to an important conclusion: the policy had on average a positive and statistically significant impact on several dimensions of academic performance. Specifically, students affected by the change in merit requirements earned more credits in the first year of enrollment and improved their longer-term academic outcomes, such as on-time BA completion. This is a very important result, given that in Italy many students delay graduation because the system allows them to get better grades while staying longer in HE (specifically, by retaking exams until getting the desired grade, see Section 2). However, the effects of increasing merit requirements are heterogeneous. We find positive and significant effects on first-year GPA, credits earned and on-time graduation (and with at most on-year delay) for medium and higher-ability students and negative effects on persistence into the second year and final graduation marks (conditional on on-time graduation) for lower-ability students.

In this paper, we focus on Italy, which we believe could be an interesting case for several reasons. First, tuition fees at universities in Italy are relatively high compared to other European countries (European Commission, 2015) and have been rising since the introduction of the "Bologna Process," particularly after the economic crisis in 2008 (Civera et al., 2020). Therefore, need-based financial aid constitutes an important part of Italian HE to ensure that disadvantaged students are not discouraged from entering HE. Second, the Italian HE has long been characterized by a very high proportion of dropouts and graduates behind schedule. Significant numbers of dropouts occur during the first year of enrollment (Mealli and Rampichini, 2012), and this seems to be explained by university costs (tuition fees and housing costs), financial constraints, and lack of access to need-based aid (Modena et al., 2020; Ghignoni, 2017). Third, within the Italian context, PoliMi is an even more intriguing case. It is a prestigious, selective, public university (ranked #1 for Engineering in Italy, and among the top 20 in Europe) whose academic offerings are focused on STEM majors that yield higher earnings returns than other majors (Altonji et al., 2014). The implications of our analysis are discussed in the conclusions.

This paper makes important contributions toward designing financial aid packages. Results suggest that student aid design affects student academic performance. Our analysis reveals highly heterogeneous effects by students' levels of ability (or academic preparedness as reflected in the admission test scores), showing that increasing efficiency may come at the cost of also increasing inequality in HE. Indeed, we find that the reform reduced persistence into the second year for lower-ability low-income students. Although at first glance, this may seem consistent with the aim of the reform, i.e. reducing student aid to "underperforming" students, it nonetheless raises serious concerns. As for efficiency, it must be kept in mind that by "lower-ability" students, here we are referring to students who still managed to enter a very prestigious and highly selective institution through an admission test. Students ranked as lower ability at PoliMi may be on the right tail of the ability distribution in less selective HE institutions. Thus, the negative effects on second-year enrollment for lower-ability low-income students may be exacerbated at less selective universities adopting similar policies. Moreover, student dropout may have negative effects on the supply of STEM graduates, which are highly demanded in the labor market. As for equity, our analysis suggests that the reform might have exacerbated inequality in HE. Unlike those receiving aid, lower-ability students from advantaged family backgrounds (i.e. not receiving financial aid) were not negatively affected by the reform. This raises the question of where HE institutions should place the bar of merit requirements. Overall, the main message is that effectively balancing the need and merit requirements of need-based programs could be the key to success in guaranteeing equality of opportunities for disadvantaged students while increasing efficiency in student aid. One way of doing it could be replacing the existing "*all or nothing*" financial aid packages with *"sliding scale"* packages, in which students only partially lose financial aid depending on their performance.

This paper is organized as follows. Section 2 describes the institutional background, namely the financial aid system in Italy. Section 3 reviews the literature about the effects of merit-based financial aid on student performance in Europe and the US. Section 4

provides a simple conceptual framework to discuss the effects of financial aid packages and interpret the empirical results. Section 5 presents the data and discusses identification. Section 6 present the results from our empirical analyses and some robustness checks. Section 7 discusses the main findings and concludes.

## 2. Institutional background: the Italian financial aid system, *Politecnico di Milano* and the reform of 2011/12

The Italian financial aid system is regulated at the national level and called *Diritto allo Studio Universitario* (DSU, hereafter). "DSU-aided" students are those receiving grants under the national financial aid system. The financial aid package has three components: grants, free lunch, and housing. Grants are by far the major component of financial aid; indeed, in addition to a financial transfer per year, eligible students are exempted from tuition fees.

The grants amount depends on whether the student lives in the city where the university is located (*in-site*), if she commutes to the city (*commuting*) or if she is an out-of-site student (*out-of-site*). In our period of analysis, the minimum amount of the grant was on average 2,000€, 3,000€, and 5,000€ for the three categories, respectively (every year, the amount is determined by a national government's decree, and adjusted by inflation). Grants are assigned based on financial need when students enroll for the first time in a bachelor's or master's degree. Applicants' financial need is assessed through an index that is an equivalized economic situation indicator (called ISEE), which is computed based on family income and the level and composition of wealth. The threshold for eligibility is set by the national government every year.[1]

The State Supplementary Fund represents the main line of financing to support the DSU system. Over the years, this fund has undergone significant reductions, particularly in 2010 when the funds went down from 250 million to 97 million €.[2] When national and state financial resources are not sufficient to cover all eligible students, some universities allocate their own funds to fill the gap. PoliMi allocates between 3 and 5 million € every year for this purpose, guaranteeing that all eligible students receive the grant. However, this was not possible in 2010 due to the significant reduction in state funds that led to a drop in the number of DSU beneficiaries (see online Appendix Figure A1). From 2011 onwards, PoliMi agreed with student unions to reinstate the compensation for the lack of State resources with university funds.

To maintain aid eligibility after the first year of enrollment, students need to pass a minimum number of formative credits (CFU henceforth) in a given academic year to meet merit requirements as well as maintain need-based eligibility.[3] In the first year, students apply for the grant directly to the university in June/July before the beginning of the academic year; classes start in September. Students receive provisional information about their grant eligibility in October, and then a final confirmation in December when students receive the first half of the grant. Students' academic performance is then assessed on August 10$^{th}$. If they meet the minimum number of credits required to maintain eligibility, they receive the second half of the grant – otherwise, they are not eligible anymore. If students reach the minimum required number of CFUs, they are automatically eligible for the grant in the subsequent year and receive the first half in September. Students in the first year also have an additional requirement to fulfill; those who do not meet the merit threshold before August 10$^{th}$ are scrutinized again on November 10$^{th}$. If by then they do not meet the merit requirements, they must pay back the first installment of the grant. Thus, meeting the merit requirement to maintain aid eligibility and avoid repaying grants is a high-stake incentive toward reaching satisfactory academic progression in the first year. It is worth noting that these students are only required to pay back for the first installment and not for the tuition and fees they were exempt from in the first year.

Before 2011, the national regulation identified the threshold for maintaining eligibility in the second year at 25 CFUs (out of 60), to be met before August 10$^{th}$ (before the last exam session of the first year, which are held in September). To maintain eligibility in the second and third year, students are required to earn 80 and 135 CFUs before August 10$^{th}$, respectively– so, after the first year, the students were expected to earn 55 credits (out of 60) every year. At this pace of credit accumulation, students are supposed to graduate almost within the legal time to completion (three years). In the Italian HE system, delayed graduation is quite frequent, and only 58% of BA students in 2020 completed their studies within the legal time (Almalaurea Consortium, 2021).[4]

In Spring 2011, the Lombardy Region (the most populous region of Northern Italy where PoliMi is located) introduced a change in the financial aid system, starting to operate from the academic year 2011/12 –for first-time-in-college students enrolling in September 2011. Thus, the reform would not affect students already enrolled during the academic year 2010/11.[5] The Regional government reached an agreement with local universities to increase the merit requirement for maintaining aid eligibility in the second year. Based on student performance from prior years, the regional government suggested a benchmark of 35 (out of 60) for universities to change their merit-based criteria for grants.

[1] Applications are submitted and eligibility is communicated weeks after enrolment (usually, in October; the academic year starts in September). According to Eurostudent (2018), 73% of students' financial resources come from their families, 18% from work-study, and only 8% from financial aid. Although this breakdown is different for low-income students, it suggests that liquidity constraints are not a major factor that deter eligible students from attending college for a few weeks before receiving financial aid.

[2] For more details, see the report on financing grants: https://www.corteconti.it/Download?id=561d9ffa-5d5a-4cc3-90d1-4eca385b3bb4

[3] Some students qualify for an exemption due to a qualifying medical condition, pregnancy, or other extenuating circumstances.

[4] Students who are eligible for DSU may not apply for several reasons (e.g. lack of information, concerns about sharing information on their family income, anxiety or stress to meeting merit requirements), despite HE institutions regularly informing and encouraging students (by email and their websites) to apply for financial aid.

[5] The new rule was communicated to students by universities and the regional government. This information was publicly available on universities' websites.

PoliMi set a new threshold for maintaining aid eligibility in the second year, increasing it from 25 to 35 (out of 60) university formative credits[6] - i.e. a 40% increase from the baseline - for first-year BA or master students entering the academic year 2011/2012 – and onwards.[7] The number of CFUs required to maintain the grant in the second and third year remained unchanged at 80 and 135 credits, respectively. The policy and institutional rationale of this reform was that this increase in merit-based requirements would have stimulated grant recipients to study harder and to accumulate more formative credits, possibly lowering time-to-graduation and increasing on-time graduation (i.e. within the normative time to completion, three years). All the other rules remained virtually unchanged: the income threshold for obtaining the grant in the first year and maintaining it over time, the amount of the grant (except for the inflation adjustment), exemption from tuition fees, etc.

This paper aims to estimate the effect of increasing the merit requirement of the DSU program. Such reform was equivalent to passing an additional course in the first year (most courses are 10 credits in the first year, with a teaching load of 100 classroom hours) to meet the DSU renewal requirements. The paper uses two cohorts of students before (academic years from 2008/09 and 2009/10) and two cohorts after the reform (the 2011/12 and 2012/13 academic years). We restricted the sample to entering cohorts that did not face the DSU coverage gap and excluded the 2010 cohort as grant eligible students had significantly lower chances to receive DSU relative to other cohorts, due to the unexpected and substantial drop in State funding that year.

The grading and examination system in the Italian HE institutions is also part of the institutional context of this reform, given that the reform potentially impacted average GPA (in addition to the number of formative credits). Students can opt to sit exams at the end of each course and achieve passing grades between 18 and 30. If the grade earned is less than 18 (i.e. fail), students must retake the exam. However, students can also opt for "refusing" their grades (and credits earned, if passed) and retake the exams later in the next semester or even in subsequent semesters. Students can retake exams several times, delaying their graduation until they are satisfied with their GPA. At PoliMi, there are five exam sessions per year – two in the winter term, two in the summer, and one in September. No precise statistics are available about the number of exams attempted per student, however retaking exams after refusing grades is quite usual.[8] The final graduation mark ranges between 66 and 110 and is computed by summing up a starting mark that depends on GPAs and an evaluation of the final exam (usually consisting of a written dissertation).

In the Italian context, GPA is therefore somehow not only driven by student effort and ability, but also by student choice (at least partially). Students put a lot of weight on their GPA, and they may influence to some extent their academic careers and time-to-degree by choosing the timing and the order of exam taking. Students can take exams several times, which gives them an incentive to cherry pick their grades to improve their GPAs at the cost of delaying graduation. Although some Italian universities (e.g. the University of Bologna, UNIBO) have tried to introduce grading reforms to limit grade refusals, these efforts failed following students' protests. Students argued that they would have been penalized by this reform, especially those who needed high grades to enroll in specialization courses. Italian students perceive GPA as "signals'' in the labor market and are particularly interested in a high GPA and final graduation marks because grades are not only important for graduate school admission or specialization courses but are also an eligibility criterion to access high-level public sector jobs (e.g., a final graduation mark of 105 out of 110 is often required). As a result, some students may choose to prolong their studies to achieve better grades, or simply because they put more weight on leisure rather than on studying. Yet, delayed graduation generates concerns as it postpones entry into the labor market, which diminishes labor supply and may induce earnings loss (both as opportunity costs and at labor market entry). Prior research found negative effects of delayed graduation on employment and earnings (Aina and Pastore, 2020, Aina and Casalone, 2020; Piazzalunga, 2018; Aina and Casalone, 2011). By contrast, some studies demonstrated that in Italy a higher final graduation mark reduces the risk of overeducation (Argentin, 2010) and is associated with higher earnings (Piazzalunga, 2018).

The reform we are studying encourages lower time-to-degree by implicitly rewarding it through maintaining student aid and by stimulating high performance in the first year. Thus, this reform might have changed the relative weight students put on higher GPAs in favor of lowering time-to-degree.

## 3. Related literature

Student aid design matters and varies across different types of students. Dynarski and Scott-Clayton (2013) provide an extensive review of the effectiveness of financial aid programs in the US and suggest that design matters for improving student performance. Specifically, their study suggests that merit-based incentives within the grant/aid systems help stimulate better performance of eligible students.

The impact of merit-based financial incentives on student performance has been previously studied in the US and European contexts. Scott-Clayton (2011) conducted a rigorous evaluation of the role of merit-based incentives by examining the PROMISE program in West Virginia, which offers free tuition to students who maintain a minimum GPA and course load. Using a regression

[6] One CFU generally corresponds to 25 hours of student workload.

[7] In the baseline models, we estimate the average effect of the reform for all majors, although the general rule described here applies to Engineering that are the most popular majors (see Table 1, about 60% of non-DSU and 65% of DSU students before the reform), while differences exist for Architecture and Design majors. For Architecture, the number of credits required was already higher and equal to 30, so the reform only increased the number of credits by 5. For Design majors, PoliMi set an even higher threshold of credits– i.e. 40 instead of 35 (after the reform). Thus, the treatment intensity was largest for Design and lowest for Architecture degrees. As a result, there will presumably be no single effect of the reform on student academic outcomes but rather differential effects by major.

[8] One limitation of the data is that we do have information of exams passed but refused or exams taken but failed.

discontinuity design (RDD), she found that this financial incentive program had significant positive effects on academic achievement. By exploring the mechanisms behind her results, the author found that the effects were concentrated around the annual merit requirements for scholarship renewal and not around financial constraints.

Jones et al. (2022) studied a change in the rules to receive full-tuition scholarships provided to high-achieving students through Georgia's HOPE Scholarship program. Under the new rules, only students meeting more rigorous merit-based criteria would retain the original scholarship covering full tuition, now called Zell Miller, with other students seeing aid reductions. They exploited discontinuities around high school GPA and SAT (Scholastic Assessment Test)/ACT (American College Testing) cutoffs for Zell Miller eligibility to analyze the effect of partially losing financial aid on college persistence and graduation of high-achieving students. The authors did not find evidence that the financial aid reduction affected persistence or graduation for these students, which suggests that high-achieving students may be less sensitive to prices once they are in college.

Belot et al. (2007) drew upon a major reform in the financial aid system in the Netherlands to study the impact of student support on the performance and time allocation of students. The reform, which was introduced in 1996, reduced the maximum duration of grants by one year and limited it to the nominal duration of studies. The authors employed a DID approach to assess the impact of this reform using cohorts of students between 1995 and 1997 (one year before and one year after the reform). They found that reform had positive effects on student performance in the first years of HE (students switched less to other programs and obtained higher grades) but did not have an impact on the time allocation of students.

The role of merit-based financial incentives was explored in a randomized trial conducted at the University of Amsterdam in 2001/02. A sample of first-year students was eligible to receive a cash amount if they passed all their exams. The results of the experiment, reported by Leuven et al. (2010), revealed that the incentive had a positive effect on high-ability students and a negative effect on low ability students.[9] This study, albeit giving useful insights on the potential role of merit-based components of financial interventions, cannot directly inform design principles of financial aid offers as the intervention was not a financial aid program, but instead a "pure incentive" (i.e. a merit-based incentive).

While the literature investigating the effect of introducing merit-based requirements on student performance offers several important contributions, evidence of the effects of tightening the merit component of student financial aid is scant. To the best of our knowledge, Montalbán (2019) is the only paper to examine this effect in the context of the *Becas de Carácter General* (BCG, the Spanish national financial aid program for low-income students in tertiary education) at the Carlos III University. Leveraging a financial aid reform that changed the minimum merit requirements for both credits earned and GPA, the author compares students who only received a fee waiver with those who were eligible and also received a cash allowance. Montalbán found that the latter performed significantly better than the former in terms of GPA and credits earned, but only after the reform. By contrast, he did not find significant positive effects when comparing students who received different amounts of cash allowances and concludes that student aid design rather minimum merit requirements may improve student performance.

Our analysis has some major differences with Montalbán (2019). First, he focuses only on *students who applied for student aid* and received some form of aid (at least a fee waiver) and uses a Regression Discontinuity Design (RDD) with different income thresholds, which determine different types of treatments within the population of students receiving aid. In his case, all students receiving grants were subjected to the same increase in merit requirements after the reform. In our paper, identification is based on a DID design in which we compare changes over time in student performance for those students who were affected by the increase in credit requirements after the reform, with those who were not receiving aid and therefore were not affected by the reform. Second, DID allows us to estimate effects that are less "local" than those provided by Montalbán (2019), which apply to students in the vicinity of the income thresholds determining the different types of treatments. Another difference is related to the characteristics of the BCG aid reform that allow Montalbán to isolate the effects of increasing merit requirements. He estimated the effect of receiving different amounts of student aid (a fee waiver vs. a low-grant amount) under "weak" and "strong merit" requirements. The fee waiver and the low-grant amount however only apply to relatively higher-income students among those receiving aid. In contrast, we estimate the average effect of increasing merit-requirements for all students receiving aid irrespective of their income levels (and therefore of the amount of aid received).

Finally, there are institutional differences between Italy's and Spain's university and student aid systems. Undergraduate degrees in Spain last four years while in Italy three years. In the case analyzed by Montalbán, students could receive aid up to one year (two years for STEM degrees) after the official length of the degree, while in Italy only for the official length of the degree. Finally, the increase in merit requirements in the case of Montalbán not only included a much larger increase in the number of credits to be earned but also a new GPA requirement and a new formula to compute the amount of student aid. Overall, the BCG reform had a greater performance-incentivizing component than the reform studied in this paper. The reform we examine was much simpler and only introduced a higher credit requirement. Our paper provides direct evidence on the effect of increasing merit requirements and complements previous evidence of similar reforms that had a different design and were implemented in different HE contexts.

Our paper is also related to recent work by Scott-Clayton and Schudde (2020), who investigate the effect of failing satisfactory academic progress (SAP, hereafter) requirements on student performance in one state in the US. The authors found that failing SAP has large discouragement effects. Their preferred DID estimates point to a decrease in second-year enrollment of 6 percentage points, but an increase of GPA of 0.11 grade points for students close to the merit threshold. Moreover, longer-term effects are negative, with

[9] A similar experiment conducted in a Canadian campus led to quite different results. The incentive was designed to offer cash incentives for course grades above 70, i.e. to stimulate a strong merit level. The results reported by Angrist et al. (2014) indicate no strong effects of this incentive scheme, with a null effect on GPA.

reductions in credits attempted and credits earned of 3.4 and 1.4 credits, respectively. This study also sheds light on the potential redistributive effects of failing merit requirements as the authors also found that dropout effects of SAP failure are stronger for low-income students. These results confirm their earlier findings from examining a SAP policy in a different state (Schudde and Scott-Clayton, 2016). Scott-Clayton and Schudde (2020) used a variety of identification strategies to estimate these effects. In an RDD, they estimate the effects of failing SAP within the population of student aid receivers. They also compare the effect of failing SAP for aided and non-aided students in a DID-RDD design. Finally, they use a DID, which enables them to estimate effects also far from the GPA cutoff of failing SAP. This last specification is the most similar to ours, as it compares aided versus non-aided students, although our work differs from Scott-Clayton and Schudde's study. While their study investigates the consequences of failing the GPA requirements, we compare two regimes, leveraging a reform that increased merit requirements. Hence, we focus on a different policy parameter. Moreover, we argue that merit requirements (and a reform increasing them) may have other effects on student performance in addition to those mediated by a higher probability of failing these requirements. For instance, in the specific case of Italy, the increased pressure towards meeting the credits requirement may reduce student GPA (in Italy students can retake exams and "accept" grades only if they are satisfied with their exam results) and induce some students to switch to another degree or institution, or even to drop out from HE. This effect may occur even for students who manage to meet the merit requirements. In other words, an increase in merit requirements may have wider effects than those mediated by failing the required performance standard.

## 4. A simple conceptual framework for the analysis of stricter merit requirements

The effect of tightening the merit requirements in need-based programs can be analyzed using a simple framework of student effort and academic performance.

Let us consider a static model in which an individual maximizes a utility function $U(C,x;a)$, where $C$ is life-time consumption of a composite good and $x$ is "student effort" (with $x \in [0,1]$), which gives disutility (i.e. a *bad*). We omit the individual subscript to simplify notation. Moreover, we assume that the disutility of studying depends negatively on individual ability $a$, i.e. time devoted to studying creates a higher disutility to lower-ability individuals. Consumption is equal to individual non-labor income $y$ (e.g. parental transfers) plus labor income, with the life-time labor market reward of student effort (as a university graduate) assumed to be linear in effort, and a scholarship $s$ that is paid only for levels of performance—coinciding with effort for simplicity[10]—above a given threshold. The budget constraint then reads as $C = \alpha x + y + s\ I(x \geq x_0)$, where $I(.)$ is the indicator function. The scholarship $s$ can be thought of as additional lifetime "study" income that is provided to the student if performance (effort in this simple model) is above a given cutoff $x_0$. The reform consists in increasing the level of performance above which the scholarship is paid from $x_0$ to $x_1$.

First, we must distinguish between eligible and ineligible students for student aid according to family income.[11] For the second group, the change in merit requirements does not change individual incentives and should not produce any effect on student performance.[12] So in our empirical analysis, non-aided students will make the natural control group. Conversely, the behavior of students receiving financial aid is likely to be impacted by the reform. The increase in the merit requirements, in the absence of an increase in the grant amount, cannot have a positive effect on individual utility because as shown in panel a Figure A2 (online Appendix) the set of feasible consumption bundles is reduced by the grey area $F$ after the reform. Thus, in this simple framework individual utility either remains the same or is reduced after the reform. However, ex-ante we cannot say anything on the direction of the change in the optimal amount of effort $x^*$, which may increase, decrease, or remain the same as before the reform, with effects on academic performance that go in the same direction.

Figure A2 shows five examples of possible effects of the reform. Subfigure (a) depicts the case of a student who before the reform was choosing a low level of effort (and consumption) and is not affected by the reform (i.e. his or her pre-reform optimal bundle is still feasible after the reform). This is likely to happen when indifference curves start to be steep at very low levels of effort, e.g. for lower-ability students who require a high level of consumption to compensate for even a relatively small increase of effort. Subfigure (b) displays the situation in which the lower-ability student (steep indifference curves) chooses a higher level of effort after the reform compared to before: academic performance improves, but utility decreases. In particular, even if in the student's optimal choice the student raises effort and performance, he or she loses the scholarship as performance is below the new threshold $x_1$. Subfigure (c) depicts the case in which the reform induces a lower-ability student to reduce effort and performance and lose the scholarship. Subfigure (d) shows the situation in which the student was choosing a very high level of effort before the reform, e.g. because indifference curves are very flat, i.e. a higher-ability student. In this case, the student is not affected by the increase in merit requirements. Other cases are also possible, for instance, some students (most likely students with an intermediate ability) might increase effort and performance and retain the scholarship after the reform, as shown in Subfigure (e).

[10] For ease of exposition, we assume that labor income is linear in academic performance and that the educational production (or performance) function is *Performance* = *f(x)* = $x$, i.e. only effort matters. Assuming a more general production function *Performance* = *f(x,a)*, for instance, would make the budget constraint nonlinear in effort and depend on an individual's ability without changing the main implications of the sketched model. In the model we remain agnostic on the meaning of "student performance", it may refer to the GPA or the number of credits achieved.

[11] For instance, in our empirical framework the budget constraint is $C = \alpha x + y + s\ I(x \geq x_0)$, only if $y \leq y_0$, and $C = \alpha x + y$ if $y > y_0$, that is only students with family income (transfers) below a given cutoff receive financial aid.

[12] To be more precise, in the case of spillovers from the treated to the untreated students (e.g. peer group effects), the performance of the untreated might also be affected. Given the low proportion of students receiving grants in Italy, and particularly at PoliMi (3%), spillover effects are unlikely to be a major concern in our case compared to other HE systems where the coverage of student aid is more substantial.

This simple representation of a student's choice problem suggests that the effect of the reform can be *heterogeneous* and depend, among other things, on the level of family income ($y$), the returns to education ($\alpha$), the amount of student aid ($s$), how much the merit-based requirements are increased ($x_1 - x_0$) and how indifference curves are shaped with respect to student ability. Thus, when assessing the issue empirically, it is important to focus not only on the average effect of the reform but also to investigate heterogeneity across different student groups (by level of ability, for instance). As for the interplay between the reform and student ability, for instance, according to the conceptual framework above, we put forward that *the reform is more likely to have positively affected the academic performance of intermediate-ability students and negatively affected that of lower ability, while leaving unaltered the performance of higher-ability students*.

In a fully-fledged dynamic model in which students decide whether to enroll in HE or not and academic performance is not fully deterministic but subject to random shocks, students can decide whether to continue or to drop out from college and we might expect additional effects of tightening the merit requirements of student aid. For instance, some students (most likely the lower ability ones) may fail to meet the requirements and decide to drop out from HE. Moreover, the reform may modify the average ability of entrants in HE (*sorting*), especially that of low income (who receive DSU), since the higher effort to maintain the grant after the reform would entail a lower increase in disutility for higher-ability students compared to their lower-ability peers.

## 5. Data and identification

### 5.1. Data

This paper uses administrative data from PoliMi.[13] PoliMi's student data-tracking system gathers information on students' background characteristics (including standardized admission test scores taken during secondary school), financial aid, and transcript and degree information such as year-by-year college enrollment (e.g., CFUs, GPA). One thing the financial data do not include before 2013 is the amount of financial aid and income eligibility measures (e.g., ISEE). However, we observe an indicator for receiving financial aid and tuition fee levels, and we use this last piece of information in the empirical model as a proxy of socio-economic background (as explained in the subsequent sections).

We focus on first-time-in-college BA students who entered PoliMi between the 2008/2009 and 2013/2014 academic years, and we follow these cohorts until 2017/2018. We limit the sample to students who entered PoliMi when they were between 17 and 25 years of age and to Italian students. We further restricted the sample to entering cohorts that did not face the DSU coverage gap of 2010 and excluded the 2010 cohort as eligible students had significantly lower chances to receive DSU relative to other cohorts. Our results are robust to the inclusion of the 2010 cohort (see online Appendix Table A5). Across these six student cohorts, we have about 27,000 observations.

For each student, we distinguish academic outcomes between short-term and long-term outcomes: (i) short-term academic outcomes include credits earned by August 10$^{th}$ of the first year (first deadline to meet merit requirements), credits earned by November 30$^{th}$ (second deadline to meet merit requirements), credits earned in the first year, GPA on August 10$^{th}$ of the first year, GPA on November 30$^{th}$ of the first year, first-year GPA, probability of enrolling in the second year; and (ii) long-term academic outcomes including the probability of enrolling in the third year, the probability of on-time graduation or with at most one-year delay, and the final graduation mark (conditional on graduation).

Table 1 displays descriptive statistics separately for the periods before and after the change in the merit component of the DSU financial aid program. Almost 3% of the sample received DSU grants during their first year of enrollment—we label them "DSU-aided" (or simply DSU) students. This is a small percentage of coverage compared with those reported for student aid programs in other countries.[14] Thus the coverage of 3% observed at PoliMi is low even by Italian standards. The average percentage of Italian students receiving DSU grants irrespective of enrollment year and type of degree (pooling undergraduate and postgraduate degrees, excluding Ph.D. degrees) was about 6.7%.[15] Consistent with the highly selective nature of PoliMi's degrees, low-income students are less represented relative to the Italian HE system.

DSU students are less likely to be female. Moreover, we observe a reduction in female DSU students after the reform. DSU students have higher admission scores compared to non-DSU and the gap has increased after the reform. Given that students are aware of the rules for student aid in advance, tightening the merit-based component may produce "cream-skimming" of low-income students, attracting those with relatively higher ability. Indeed, Table 1 shows a relative increase in DSU students' admission test scores after the reform. In the empirical analysis, we discuss and address this concern.

Table 1 hints at some positive effects of the reform on student performance. The number of credits earned in the first year, the first-year GPA, and the probability of on-time graduation increased after the reform. However, these differences may hide changes in student cohort characteristics. For this reason, we implement a DID estimation strategy.

[13] We are grateful to the Offices for Student Support and for Information Systems for sharing the data and helping us clean and understand the variables in the dataset.

[14] In the U.S. context, Dynarski (2003) reports a 12% coverage at the peak of the Social Security Student Benefit Program, and Schudde and Scott-Clayton (2016) report a coverage of at least 30% for the Federal Pell grant. Montalbán (2019) reports that about a quarter of undergraduate students were covered by the Becas de Carácter General need-based grant program. In general, the percentage of students receiving aid in Italy is lower compared to other EU countries of similar population size (European Commission, 2015).

[15] Authors' estimations using publicly available data from the Ministry of University and Research.

**Table 1**
Means of student characteristics and student outcomes by DSU aid status, before vs. after the reform

| | non-DSU | | Diff. non-DSU | DSU | | Diff. DSU | DID | DID |
|---|---|---|---|---|---|---|---|---|
| | a | b | (b-a) | c | d | (d-c) | (d-c)-(b-a) | S.E. |
| | Before | After | | Before | After | | | |
| Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| **Student outcomes** | | | | | | | | |
| *Short term* | | | | | | | | |
| Credits earned 1st year | 39.346 | 40.136 | 0.79*** | 50.360 | 53.511 | 3.151*** | 2.361*** | 0.789 |
| Credits earned Aug 10 | 34.219 | 34.358 | 0.139 | 45.331 | 49.253 | 3.922*** | 3.783*** | 0.643 |
| Credits earned Nov 30 | 39.932 | 40.271 | 0.340 | 50.434 | 53.603 | 3.169*** | 2.83*** | 0.785 |
| GPA 1st year | 24.068 | 24.055 | -0.013 | 24.503 | 24.837 | 0.334* | 0.347* | 0.184 |
| GPA Aug 10 | 24.195 | 24.182 | -0.013 | 24.566 | 24.919 | 0.352* | 0.365* | 0.187 |
| GPA Nov 30 | 24.076 | 24.057 | -0.020 | 24.508 | 24.843 | 0.335* | 0.354* | 0.184 |
| Enrolled in the 2nd year | 0.806 | 0.799 | -0.007 | 0.975 | 0.960 | -0.016 | -0.008 | 0.013 |
| *Long term* | | | | | | | | |
| Enrolled in the 3rd year | 0.704 | 0.721 | 0.017*** | 0.909 | 0.960 | 0.05*** | 0.033* | 0.018 |
| Graduated in 3yrs | 0.320 | 0.331 | 0.011* | 0.430 | 0.537 | 0.107*** | 0.096*** | 0.035 |
| Graduated in 3/4 yrs | 0.577 | 0.582 | 0.005 | 0.741 | 0.825 | 0.084*** | 0.079*** | 0.029 |
| Final mark 3 yrs | 100.393 | 100.747 | 0.354** | 101.167 | 101.230 | 0.062 | -0.291 | 0.71 |
| Final mark 3/4 yrs | 97.741 | 97.991 | 0.25** | 98.636 | 98.509 | -0.127 | -0.378 | 0.624 |
| N. | 12,046 | 13,809 | | 486 | 348 | | | |

| | non-DSU | | Diff. non-DSU | DSU | | Diff. DSU | DID | DID |
|---|---|---|---|---|---|---|---|---|
| | a | b | (b-a) | c | d | (d-c) | (d-c)-(b-a) | S.E. |
| | Before | After | | Before | After | | | |
| Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| **Student characteristics** | | | | | | | | |
| Female | 0.335 | 0.318 | -0.017*** | 0.374 | 0.279 | -0.096*** | -0.078** | 0.033 |
| Admission test score | 0.058 | 0.049 | -0.008 | 0.241 | 0.39 | 0.150** | 0 .158*** | 0.061 |
| Age at entry | 19.241 | 19.305 | 0.064*** | 19.1 | 19.069 | -0.036 | -0.10*** | 0.037 |
| *Residence* | | | | | | | | |
| Milan | 0.136 | 0.118 | -0.018*** | 0.064 | 0.049 | -0.015 | 0.003 | 0.017 |
| Lombardy | 0.611 | 0.424 | -0.012** | 0.599 | 0.325 | -0.099*** | -0.087** | 0.034 |
| Other North-West | 0.046 | 0.056 | 0.002 | 0.048 | 0.029 | -0.027* | -0.029** | 0.014 |
| North-East | 0.1 | 0.082 | 0.011*** | 0.111 | 0.109 | 0.027 | 0.016 | 0.021 |
| Centre | 0.024 | 0.049 | 0.009*** | 0.034 | 0.08 | 0.031* | 0.022 | 0.018 |
| South | 0.048 | 0.058 | 0.01*** | 0.237 | 0.319 | 0.082*** | 0.072** | 0.032 |
| Islands | 0.02 | 0.072 | 0.003* | 0.023 | 0.078 | 0.006 | 0 .003 | 0.019 |
| Missing residence | 0.014 | 0.009 | -0.005*** | 0.016 | 0.011 | -0.005 | 0.000 | 0.008 |
| *College major* | | | | | | | | |
| Engineering | 0.59 | 0.571 | -0.019*** | 0.652 | 0.73 | 0.078** | 0.097*** | 0.033 |
| Design | 0.149 | 0.126 | -0.023*** | 0.169 | 0.106 | -0.062*** | 0.039 | 0.037 |
| Architecture | 0.215 | 0.206 | -0.009 | 0.146 | 0.101 | -0.046** | -0.037 | 0.024 |
| Other | 0.046 | 0.097 | 0.051*** | 0.033 | 0.063 | 0.03** | -0.021 | 0.016 |
| *Fee brackets* | | | | | | | | |
| 1 | 0 | 0 | 0 | 0.994 | 0.989 | -0.005 | -0.005 | 0.007 |
| 2 | 0.108 | 0.125 | 0.017*** | 0.006 | 0.011 | 0.005 | -0.012 | 0.008 |
| N. | 12,046 | 13,809 | | 486 | 348 | | | |

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Robust standard errors in parentheses.

*Note.* This table reports sample means of student outcomes and student characteristics of students from Milan Polytechnic for the cohorts before the reform (2008/9, 2009/10, 2010/11) and after the reform (2011/12 and 2012/13) by student aid status (DSU vs. non-DSU). Sample restricted to non-mature (i.e. aged 17-25) and Italian students. The 2010 cohort is excluded. The DID column shows the differences between DSU and non-DSU students, after vs. before the reform. Admission test scores are standardized by major and student cohort (to have zero mean and one standard deviation).

### 5.2. Identification of the causal effect of the reform

The natural setting to evaluate the effect of the reform is a DID research design. The DID approach enables us to make a comparison between DSU and non-DSU students before and after the reform that tightened the merit requirements. We estimate the effect of increasing the merit requirements of the DSU program on the outcome variables described in the previous section using the following parametric specification:

$$Y_{it} = \alpha + \gamma(DSU_i * After_t) + \delta Cohort_t + \sigma DSU_i + \lambda x_{it} + u_{it} \quad (1)$$

where $Y_{it}$ represents the outcome of student $i$ of cohort $t$. $After_t$ is a dummy variable that takes the value of zero if the student entered for the first time at PoliMi in 2008 or 2009 (pre-reform period) and one if the student entered between 2011 and 2013 (post-reform period). $DSU_i$ takes the value of one if the student received the DSU grant (the treated group) at first entry and zero if the student did not meet the income eligibility requirements to receive financial aid (the control group). $Cohort_t$ is a vector of student cohort fixed effects, $\mathbf{X_i}$ is a vector of student-level covariates that controls for demographic and pre-college ability and $u_{it}$ an idiosyncratic error term. Covariates included are gender (female), score in the standardized admission test (proxying ability), dichotomous indicators for student fee brackets (proxying income, and in general socio-economic status), residence (mutually exclusive categories: Milan – reference, Lombardy, Other North-Western regions, North-Eastern regions, Central regions, Southern regions, Insular regions, missing residence), students' area of study or major (Engineering – reference, Design, Architecture, Other). $\beta$ is the DID effect of interest given by the interaction $DSU_i$*$After_t$.

## 6. Results of the empirical analysis

### 6.1. Baseline results and robustness checks

The baseline results for the estimation of the effects of the reform on student performance are reported in Table 2. For the short-term outcomes, the reform increased the number of credits earned in the first year by 1.8, and the number of credits earned by August 10th and November 30th of the first year, by 3.2 and 2.3, respectively. These effects are statistically significant at 1% and 5% level. These figures correspond to 3%, 7%, and 4.6% increases, respectively, at the baseline (that is before the reform) for students receiving financial aid. The effects are largest for credits earned by August which is consistent with the fact that the first deadline was the most important one. Meeting the merit requirements by August entails receiving the second installment and avoiding the risk of failing the second deadline and having to repay the first installment. Some of these improvements fade away as students approach the end of the academic year, suggesting that they may optimally allocate their time and increase effort in the first part of the academic year to meet the first deadline while reducing it in the rest of the year after achieving this important goal. We also find positive effects on GPA, albeit less precisely. GPA increases by around 0.3 points, corresponding to an increase of 1.2% relative to the baseline mean.

The positive effects on student progression persist into graduation. We find positive effects on the probability of on-time graduation (within three years), which increased by about 9 percentage points (pp, hereafter), corresponding to a 21% increase, and on the probability of graduating in three or four years (i.e. with at most one-year delay), which increased by 7.6 pp (10%). Overall, the reform appears to reduce time-to-degree which is an important finding given the large and persistent share of Italian students who spend more years than the normative time to attain a degree (Aina and Pastore, 2020) and the high percentage of students delaying graduation (Aina et al. 2011, Garibaldi et al. 2012).[16]

Next, we present robustness checks on the baseline specification. First, we check the sensitivity to controlling for a second-degree polynomial in student admission test score -an important predictor of student outcomes. This is an important check as the reform might have changed the ability profile of DSU students enrolling in PoliMi, thus affecting their performance. The results in Table 3 are reassuring. The estimated effects are very close to those in Table 2, namely 1.9, 3.2, and 2.4 for credits earned in the first year, August 10$^{th}$ and November 30$^{th}$, respectively. Long-term outcomes are also very close, 8.8 pp and 7.6 pp on the probability of on-time graduation or with at most one year of delay, respectively.

The baseline estimates in Table 2 include students from all family income backgrounds. However, students receiving aid come from low-income households. Thus, we might have a lack of common support between DSU and non-DSU students, and this may be problematic especially if family income has a strong independent effect on student performance over and above student aid. This motivates the robustness check we implement in Table 4, which shows the estimates of equation (1) in different samples that gradually include students with higher family income. We compare the estimates when including students in the first four, first five, first six, and all tuition fee brackets. Two main results stand out in Table 4. First, estimates on credits earned are quite robust to varying the sample composition by family income. Although estimates on the number of credits earned in the first year and on August 10$^{th}$ and November

[16] It is worth noting that in the first year of enrollment most courses (if not all) are compulsory, and students have few degrees of freedom (often none) to take elective courses. In the second and third years, they have somewhat limited choice in creating their study plan. Most importantly, students must create their study plan before taking exams and cannot include exams in which they had the best performance. In the first year, students might respond to the higher merit requirements by taking easier exams to meet the credit cutoff by August. Using median grades by major and course, we identified "difficult" courses taken in the first year as those that are in the lowest quartile of the grade distribution. The reform significantly increased the number of credits earned in "difficult" courses by 2, suggesting that course choices are unlikely to explain our findings on CFUs. The reform seems to have positively affected first-year student performance by increasing student effort instead of choosing easier courses.

**Table 2**
DID estimates of the effect of the reform (Baseline estimates)

| Short-term outcomes | Coeff. | S.E. | N. | Long-term outcomes | Coeff. | S.E. | N. |
|---|---|---|---|---|---|---|---|
| Credits earned 1st year | 1.827** | (0.832) | 26,689 | Enrolled in the 3rd year | 0.029 | (0.019) | 26,689 |
| Credits earned Aug 10 | 3.191*** | (0.850) | 26,689 | Graduated in 3yrs | 0.090*** | (0.034) | 26,689 |
| Credits earned Nov 30 | 2.332*** | (0.830) | 26,689 | Graduated in 3/4 yrs | 0.076*** | (0.029) | 26,689 |
| GPA 1st year | 0.284* | (0.165) | 23,583 | Final mark 3 yrs | -0.231 | (0.652) | 8,829 |
| GPA Aug 10 | 0.302* | (0.167) | 23,346 | Final mark 3/4 yrs | -0.350 | (0.560) | 15,636 |
| GPA Nov 30 | 0.292* | (0.165) | 23,585 | | | | |
| Enrolled in the 2nd year | -0.020 | (0.013) | 26,689 | | | | |
| Control variables | Yes | | | | Yes | | |

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses.
*Note.* The Table shows the DID estimates (i.e. the $DSU_i$*$After_t$ coefficient) of equation (1). Post-reform cohorts (After) are those enrolled after 2010/2011. The number of observations varies across student outcomes as GPA is only available for exams that attribute a grade ("pass" or "fail" grades are excluded). Among the long-term outcomes, graduation marks are conditional on graduation. The 2010 cohort is excluded. Control variables include gender, age, entry test score, income brackets, residence and college major. N. is the number of observations.

**Table 3**
DID estimates of the effect of the reform (with a second-degree polynomial in student entry test scores)

| Short-term outcomes | Coeff. | S.E. | N. | Long-term outcomes | Coeff. | S.E. | N. |
|---|---|---|---|---|---|---|---|
| Credits earned 1st year | 1.906** | (0.822) | 26,689 | Enrolled in the 3rd year | 0.030 | (0.018) | 26,689 |
| Credits earned Aug 10 | 3.235*** | (0.845) | 26,689 | Graduated in 3yrs | 0.088*** | (0.034) | 26,689 |
| Credits earned Nov 30 | 2.411*** | (0.819) | 26,689 | Graduated in 3/4 yrs | 0.076*** | (0.029) | 26,689 |
| GPA 1st year | 0.283* | (0.164) | 23,583 | Final mark 3 yrs | -0.227 | (0.652) | 8,829 |
| GPA Aug 10 | 0.302* | (0.167) | 23,346 | Final mark 3/4 yrs | -0.335 | (0.561) | 15,636 |
| GPA Nov 30 | 0.291* | (0.164) | 23,585 | | | | |
| Enrolled in the 2nd year | -0.017 | (0.013) | 26,689 | | | | |
| Control variables | Yes | | | | Yes | | |

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses.
*Note.* The Table shows the DID estimates (i.e. the $DSU_i$*$After_t$ coefficient) of equation (1). Post-reform cohorts (After) are those enrolled after 2010/2011. N. is the number of observations. Unlike in Table 3, the entry test score enters with a second-degree polynomial instead of linearly. The number of observations varies across student outcomes as GPA is only available for exams that attribute a grade ("pass" or "fail" grades are excluded). Among the long-term outcomes, graduation marks are conditional on graduation. The 2010 cohort is excluded. Control variables include gender, age, entry test score, income brackets, residence and college major. N. is the number of observations.

30th are more precise in larger samples, they are consistently of similar magnitude. Estimates on long-term outcomes follow a similar pattern and are generally larger and more precise when we consider larger samples but are of comparable magnitudes across samples. The results confirm a positive effect of the reform on both on-time graduation or with at most one year of delay.[17]

We checked the sensitivity of our estimates to omitting each of the five cohorts at a time from our sample. The results are reported in Figures A3 and A4 in the online Appendix. Given the relatively small number of treated individuals, estimates slightly change from year to year and are not perfectly aligned with the baseline estimates (labeled as "2010" [the estimates excluding that cohort] in the graphs). Yet, results generally lead to the same conclusions. Effects on credits earned and graduation outcomes are generally positive and significant.

### 6.2. *Identifying assumptions: direct and indirect effects of the reform*

The main identifying assumption in the DID approach is that the coefficient on the interaction term $DSU_i$*$After_t$ from Equation (1) would be zero in the absence of the reform. Pre- and post-reform cohorts of students may be different, and within each cohort, DSU and non-DSU students may be different; but nothing can be different about being a post-policy student receiving a DSU grant other than the new merit requirement introduced in 2011. In other words, trends in academic achievement of pre-reform cohorts should be good predictors of what would have happened in the absence of this financial aid reform in the post-reform period (parallel trend assumption).

[17] In the online Appendix Tables A3-A4, we investigate heterogeneity by major (see footnote 7). The bottom panel of the online Appendix Table A3 shows mostly no statistically significant effects of the reform on short-term outcomes for Design and Architecture majors. The reform seems to have increased the probability of on-time graduation for DSU students in all majors but Architecture (see bottom panel of the online Appendix Table A4). Since most students are enrolled in Engineering majors, and especially DSU students (see Table 1), the estimates for Architecture, Design, and other majors in Tables A3-A4 are likely to be quite noisy and based on a small number of observations.

**Table 4**
Sensitivity of DID estimates to changing student composition by family income (i.e. fee brackets)

| Short-term outcomes | Fee brackets | Coeff. | S.E. | N. | Long-term outcomes | Fee brackets | Coeff. | S.E. | N. |
|---|---|---|---|---|---|---|---|---|---|
| Credits earned 1st year | 4 | 1.633* | (0.939) | 6,802 | Enrolled in the 3rd year | 4 | 0.021 | (0.022) | 6,802 |
| | 5 | 1.791** | (0.893) | 9,880 | | 5 | 0.028 | (0.020) | 9,880 |
| | 6 | 1.669* | (0.858) | 14,708 | | 6 | 0.029 | (0.019) | 14,708 |
| | All | 1.827** | (0.832) | 26,689 | | All | 0.029 | (0.019) | 26,689 |
| Credits earned Aug 10 | 4 | 2.988*** | (0.939) | 6,802 | Graduated in 3yrs | 4 | 0.086** | (0.035) | 6,802 |
| | 5 | 3.123*** | (0.899) | 9,880 | | 5 | 0.086** | (0.034) | 9,880 |
| | 6 | 3.080*** | (0.869) | 14,708 | | 6 | 0.084** | (0.034) | 14,708 |
| | All | 3.191*** | (0.850) | 26,689 | | All | 0.090*** | (0.034) | 26,689 |
| Credits earned Nov 30 | 4 | 2.108** | (0.937) | 6,802 | Graduated in 3/4 yrs | 4 | 0.058* | (0.031) | 6,802 |
| | 5 | 2.316*** | (0.891) | 9,880 | | 5 | 0.067** | (0.030) | 9,880 |
| | 6 | 2.156** | (0.856) | 14,708 | | 6 | 0.068** | (0.029) | 14,708 |
| | All | 2.332*** | (0.830) | 26,689 | | All | 0.076*** | (0.029) | 26,689 |
| GPA 1st year | 4 | 0.258 | (0.175) | 6,038 | Final mark 3 yrs | 4 | -0.724 | (0.712) | 1,991 |
| | 5 | 0.319* | (0.170) | 8,796 | | 5 | -0.462 | (0.685) | 2,946 |
| | 6 | 0.298* | (0.168) | 13,208 | | 6 | -0.484 | (0.666) | 4,594 |
| | All | 0.284* | (0.165) | 23,583 | | All | -0.231 | (0.652) | 8,829 |
| GPA Aug 10 | 4 | 0.317* | (0.178) | 5,973 | Final mark 3/4 yrs | 4 | -0.498 | (0.599) | 3,730 |
| | 5 | 0.359** | (0.173) | 8,696 | | 5 | -0.321 | (0.581) | 5,505 |
| | 6 | 0.331* | (0.170) | 13,071 | | 6 | -0.494 | (0.570) | 8,433 |
| | All | 0.302* | (0.167) | 23,346 | | All | -0.350 | (0.560) | 15,636 |
| Short-term outcomes | Fee brackets | Coeff. | S.E. | N. | Long-term outcomes | Fee brackets | Coeff. | S.E. | N. |
| GPA Nov 30 | 4 | 0.265 | (0.175) | 6,038 | | | | | |
| | 5 | 0.328* | (0.170) | 8,796 | | | | | |
| | 6 | 0.306* | (0.167) | 13,208 | | | | | |
| | All | 0.292* | (0.165) | 23,585 | | | | | |
| Enrolled in the 2nd year | 4 | -0.020 | (0.016) | 6,802 | | | | | |
| | 5 | -0.011 | (0.015) | 9,880 | | | | | |
| | 6 | -0.017 | (0.014) | 14,708 | | | | | |
| | All | -0.020 | (0.013) | 26,689 | | | | | |
| Control variables | Yes | | | | | Yes | | | |

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses.
*Note.* The Table shows the DID estimates (i.e. the $DSU_i$*$After_t$ coefficient) of equation (1). Post-reform cohorts (After) are those enrolled after 2010/2011. The estimates are reported in samples including the first four, five, six and all fee brackets. The number of observations varies across student outcomes as GPA is only available for exams that attribute a grade ("pass" or "fail" grades are excluded). Among the long-term outcomes, graduation marks are conditional on graduation. The 2010 cohort is excluded. Control variables include gender, age, entry test score, income brackets, residence and college major. N. is the number of observations.

To discuss the potential role of time-varying students' characteristics, we draw from Zeldow and Hatfield (2021) and write the untreated potential outcome of individual $i$ as

$$E\left[Y_i^0(t)|D=d_i,\ X=x_{it}\right] = \alpha_0 + \alpha_1 d_i + \ \xi_t + \lambda' \mathrm{x_{it}} \tag{2}$$

where $\xi_t$ are time fixed effects, $d_i$ is an indicator for the treated group (with $d_i = 1$ for treated individuals, i.e. those receiving financial aid, and $d_i = 0$ otherwise), and $\mathrm{x_{it}}$ a vector of time-varying covariates. The intercept is allowed to be group specific, while $\lambda$ is a vector of coefficients for the covariates. We assume that $\lambda$ is constant over time given the brief time span considered in our analysis.

We define the treated outcome as $Y_i^1(t) \ = Y_i^0(t) + \gamma$, which implies that the estimated equation if we implement parametric DID is:

$$E[Y_i(t)|D=d_i,\ X=x_{it}] = \alpha_0 + \alpha_1 d_i + \ \xi_t + \lambda' \mathrm{x_{it}} + \gamma(p_t d_i) \tag{3}$$

where $p_t$ is an indicator for the post-policy period, which is multiplied by the treatment variable.

Zeldow and Hatfield (2021) discuss how the parallel trend assumption may fail when the outcomes depend on time-varying covariates. Intuitively, if we do not condition on $\mathrm{x_{it}}$ in the regression, the coefficient on $p_t d_i$ also captures the effect of treated and non-treated groups' characteristics that change over time. This is potentially the case of our non-experimental study, in which receiving student aid is not the outcome of a randomization, and the characteristics of DSU students may evolve differently over time, also depending on the characteristics of high school graduates who enroll in higher education, (or at PoliMi). Moreover, some of these characteristics may affect our outcomes of interest and could make the parallel trend assumption fail.

As shown by Zeldow and Hatfield (2021), controlling for time-varying covariates could be used to preserve the parallel trend assumption when the outcomes depend on time-varying covariates. Although the parallel trend assumption cannot be tested in the post-policy period, we can test whether our outcomes of interest were evolving similarly in the pre-policy period, conditional on time-varying covariates. As our pre-policy period is quite short, we include in the analysis the 2010 cohort to implement a test for the

pre-trends (for which unlike other years not all DSU eligible students were covered). We estimate a fake-DID for the years 2008, 2009 and 2010, assigning a fake treatment in 2009 (so the fake-treated cohorts are the 2009 and 2010). The coefficient on the $DSU_i$ *$After_t$ interaction captures whether the difference in student performance between DSU and non-DSU students (if any) was also changing over time in the pre-reform period. Table 5 reports the results of DID estimates with the "fake" policy assignment. The coefficients on $DSU_i$*$After_t$ are insignificant for both short- and long-term academic outcomes and are generally smaller than our baseline estimates.

To check the common trend assumption in the differences in outcomes before the reform, we have also estimated event-study DID models, for which DSU status is interacted with student cohorts (the reference year is 2009, and 2010 is omitted from the sample). Given the low share of DSU students, cohort-specific interactions are less precisely estimated compared to those in the baseline specification (Table 2). Overall, for all outcomes for which we find a statistically significant effect at 5% in Table 2, Figures A5 and A6 in the online Appendix show that the differences in outcomes in 2008 are close to those in 2009 (coefficients are similar and confidence intervals largely overlap). In the post-reform period, we observe an improvement in outcomes, especially in credits earned in August and November and the probability of graduating on time or with at most a one-year delay.

Zeldow and Hatfield (2021) emphasize that controlling for time-varying covariates can be problematic when these covariates are affected by the treatment and in turn affect the outcome. Indeed, in this case, the $x_{it}$ are endogenous or "bad" controls. These covariates, when included in the regression model, capture part of the effect of the treatment. Borrowing from Zeldow and Hatfield (2021), this can be shown as follows. First, the potential outcome definition is extended to the potentially endogenous control variables, by defining the two conditional mean vectors of the covariates $E(x_{it}^0|D=d)=\tau_{dt}^0$ and $E(x_{it}^1|D=d)=\tau_{dt}^1$, which are the means of the covariate in the presence and in the absence of treatment, respectively. We further assume that the treatment does not affect past values of $x_{it}$ so that $\tau_{d0}^0=\tau_{d0}^1=\tau_{d0}$.

In this case, the ATT is

$$ATT=E\left[Y^1(1)-Y^0(1)|D=1\right]=\gamma+\lambda\left(\tau_{11}^1-\tau_{11}^0\right)$$

with

$$E\left[Y^1(1)|D=1\right]=\alpha_0+\alpha_1+\xi_1+\lambda'\tau_{11}^1+\gamma$$

$$E\left[Y^0(1)|D=1\right]=\alpha_0+\alpha_1+\xi_1+\lambda'\tau_{11}^0.$$

Thus, the ATT is the usual DID coefficient on the treated dummy and post-policy interaction plus a term that depends on the difference in means in the covariates between the treated group in the presence of the treatment ($\tau_{11}^1$) and the treated group in the absence of the treatment ($\tau_{11}^0$). The latter, however, cannot be observed.

Zeldow and Hatfield (2021) compare two estimators, the "unadjusted estimator", which consists in estimating Equation (3) but without covariates, and the "adjusted estimator" derived from Equation (3).

The unadjusted estimator of the ATT is:

$$ATT_U=E[Y(1)|D=1]-[Y(0)|D=1]-\{E[Y(1)|D=0]-E[Y(0)|D=0]\}$$

$$=\gamma+\lambda'[(\tau_{11}-\tau_{01})-(\tau_{10}-\tau_{00})]=\gamma+\lambda'[(\tau_{11}-\tau_{10})-(\tau_{01}-\tau_{00})]$$

Thus, the $ATT$ and the $ATT_U$ recover different parameters. It is worth noting further restrictions (or identification assumptions) under which the two are equal. In particular, if $\tau_{01}=\tau_{00}$, i.e. in our specific application if the characteristics of non-DSU students do not change over time (pre- vs. post-policy period), and the assumption $\tau_{11}^0=\tau_{10}$ holds, i.e. a "before-after" assumption under which in

**Table 5**
DID estimates on the pre-policy period with fake treatment set at 2009

| Short-term outcomes | Coeff. | S.E. | N. | Long-term outcomes | Coeff. | S.E. | N. |
|---|---|---|---|---|---|---|---|
| Credits earned 1st year | 1.575 | (1.056) | 19,115 | Enrolled in the 3rd year | 0.031 | (0.026) | 19,115 |
| Credits earned Aug 10 | 1.375 | (1.115) | 19,115 | Graduated in 3 yrs | 0.025 | (0.041) | 19,115 |
| Credits earned Nov 30 | 1.619 | (1.043) | 19,115 | Graduated in 3 or 4 yrs | 0.007 | (0.037) | 19,115 |
| GPA 1st year | 0.013 | (0.193) | 16,828 | Final mark 3 yrs | -0.773 | (0.812) | 6,265 |
| GPA Aug 10 | -0.019 | (0.200) | 16,651 | Final mark 3 or 4 yrs | -0.412 | (0.658) | 11,225 |
| GPA Nov 30 | 0.010 | (0.193) | 16,831 | | | | |
| Enrolled in the 2nd year | 0.020 | (0.014) | 19,115 | | | | |
| Control variables | Yes | | | | Yes | | |

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses.
*Note.* This table reports the coefficient on $DSU_i$*$After_t$ on the sample including cohorts before the policy introduction (2008, 2009 and 2010) and imputing the reform to 2009. The number of observations varies across student outcomes as GPA is only available for exams that attribute a grade ("pass" or "fail" grades are excluded). Among the long-term outcomes, graduation marks are conditional on graduation. Control variables include gender, age, entry test score, income brackets, residence and college major. N. is the number of observations.

the absence of the treatment the characteristics of DSU students would have not changed over time, then $ATT_U = ATT$ and the ATT effect of interest is provided by the unadjusted estimator.

The adjusted estimator is instead $ATT_A = \gamma$, which only captures the "direct" effect of the policy, i.e. the part not mediated by $x_{it}$. We argue that the baseline estimates presented in the previous section correspond to the adjusted estimator. Depending on the sign of $\lambda'(\tau_{11}^1 - \tau_{11}^0)$, the adjusted estimator may represent an overestimate or an underestimate of the $ATT$. This estimator can be roughly interpreted as a sort of "behavioral effect," that is the effect driven by changes in students' behavior, while $\lambda'(\tau_{11}^1 - \tau_{11}^0)$ would be the indirect effect on outcomes determined by compositional changes induced by the reform. In general, given that retaining student aid becomes more difficult after the reform, we expect $\lambda'(\tau_{11}^1 - \tau_{11}^0) > 0$, i.e. DSU students becoming increasingly selected according to characteristics positively associated with academic performance, and therefore that $ATT_A < ATT$, i.e. our baseline estimates may provide a lower bound of the ATT.

In what follows we investigate the potential role of compositional changes in student characteristics, some of which might be induced by the reform in four ways. First, we present *F*-tests of changes in mean characteristics between treated and control groups and in the pre- and post-policy period. Second, we try to isolate the effect of compositional changes on the outcomes using a test proposed by Carrell et al. (2018). Third, we report the results of a matching-DID estimator, in which post-policy students are matched with those enrolled before the policy was introduced according to some observable characteristics. Finally, we compare the adjusted with the unadjusted ATT estimators, the former estimating only the direct effect of the reform and the latter the overall effect, including the one mediated by student characteristics, under the assumptions described earlier in this section.

Table 1 reports the means (and differences in means) of the covariates used in our analysis by DSU status and before versus after the reform. Two things are worth noting. First, some differences turn out to be statistically significant even if they are of very small magnitudes. For example, the change in means of age at entry for non-DSU students is about 23 days and is still statistically significant at the 1% level; for non-DSU students, the changes in the distribution of students by residence for each category are always smaller than 2 percentage points, but mostly statistically significant. A second interesting fact is that changes in non-DSU students' characteristics are mostly very small in magnitude and smaller than changes in DSU students' characteristics. For instance, the change in the average admission scores for non-DSU students is almost null. By contrast, changes in DSU students' average characteristics are sizable in some cases, and as a result, some of the DID estimates reported in column (7) are statistically significant at conventional levels. Overall, Table 1 suggests that the policy might have changed the characteristics of the treated group (DSU students).[18] This table also suggests that the assumption $\tau_{01} = \tau_{00}$ (the characteristics of non-DSU students did not change over time) may hold, and if in addition we impose the assumption $\tau_{11}^0 = \tau_{10}$ (the characteristics of DSU students would not have changed over time in the absence of the treatment), then $ATT_U = ATT$.

A simple balancing test on the means of the covariates may not be particularly informative for our study as it does not give any insight into which covariates are important for student outcomes. For example, a strong unbalance in an irrelevant covariate (i.e. with a zero coefficient in the outcome equation) would not pose a problem in our DID estimation. For this reason, we implement a test proposed by Carrell et al. (2018) to assess the potential effect of compositional changes on student outcomes, which has been recently applied in the context of student-aid by Scott-Clayton and Schudde (2020).[19] We regress the outcomes on the covariates $x_{it}$ (excluding therefore the cohort fixed effects, the treatment status and the $DSU_i$*$After_t$ interaction) and compute the predicted outcomes. As noted by Carrell et al. (2018) this measure captures a linear combination of individual characteristics, where the weights are chosen to best predict student outcomes. Then, we regress these predicted student outcomes on cohort fixed effects, treatment status and the $DSU_i$ * $After_t$ interaction. The latter captures the average difference in student outcomes that are predicted by changes over time in mean characteristics of DSU versus non-DSU students. We carry out two versions of the Carrell test.

In the first version, the coefficients on the covariates are estimated using the entire period of analysis (labeled as "Carrell"), and in the second the coefficients are only estimated using the pre-policy sample (labeled "Carrell pre"). Implicitly, this can also be considered as a balancing test on the predicted values. The point estimates and confidence intervals for our baseline estimates (from the adjusted estimator), the unadjusted estimator (i.e. without control variables), the "Carrell" and the "Carrell pre" model are reported in Figures 1 and 2. Three results stand out. First, although changes in covariates generally go in the direction of improving the outcomes of DSU relative to non-DSU students, consistent with a "cream skimming" (or selection) effect, in many cases these changes are not large enough to determine statistically significant changes in the outcomes, or their effects are much smaller than our baseline estimates. Second, and consistent with this selection effect, the unadjusted estimates are generally larger than our baseline estimates suggesting that the selection effect is positive. Third, using full-sample predictions instead of pre-policy predictions does not make a substantial difference, and the estimates in the "Carrell" and "Carrell pre" models are very similar. This suggests that the coefficients on the covariates are likely to be time invariant (i.e. our assumption that $\lambda$ is constant over time).[20]

In equation (3), we have assumed a constant-effect model. In this case, the DID estimator gives the average effect conditional on the

[18] We also perform a joint balancing test on all observable covariates by regressing $DSU_i$ on the covariates and their interaction with the post-reform indicator, $After_t$. The *F*-test for joint significance of the interaction terms is 0.31(p-value = 0.99).

[19] Unobservable characteristics may have also changed after the reform. We assume that these characteristics are well proxied by the observable characteristics we control for our specification.

[20] In the Carrell et al's tests, we have used White robust standard errors to account for the fact that the dependent variable was estimated from a regression (see Lewis and Linzer, 2005). As an alternative, in Figures A7 and A8 (online Appendix), we report the same graphs but using weighted least squares (WLS) using the inverse of the standard errors of the predictions as weights.

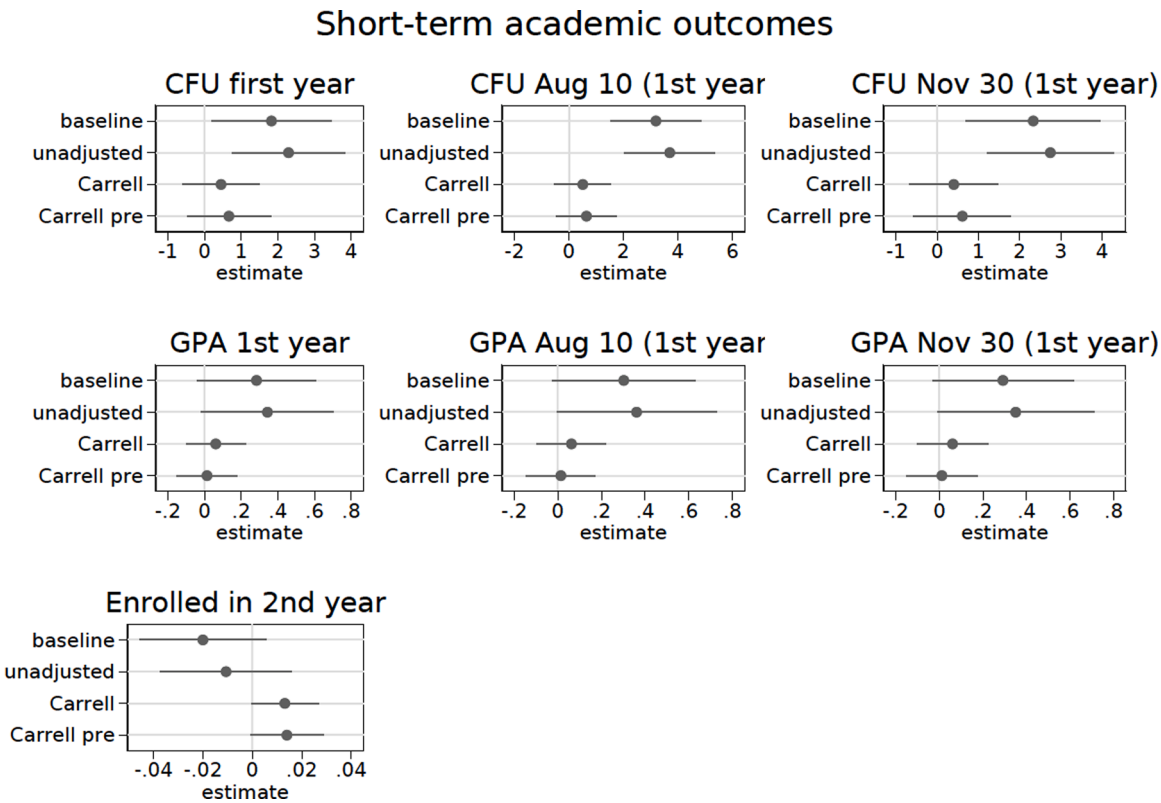


**Fig. 1.** The effect of time-varying covariates – Short-term outcomes *Note.* Each graph reports for each dependent variable the coefficient on the $DSU_i$ $*After_t$ interaction estimated on the baseline model ("baseline") controlling for the variables in Table 2, the unadjusted model (i.e. omitting the covariates, "unadjusted") and using the predictions of the student outcomes based on the covariates estimated on the whole period ("Carrell") or only in the pre-policy period ("Carrell pre"). The 2010 cohort is excluded. For all models confidence intervals are computed using White robust standard errors.

observables in the sample. In particular, the heterogeneous effects are averaged across the post-policy DSU students' characteristics, i.e. those enrolled after the reform. This however conflates the behavioral (or direct) effects with the "cream-skimming" (or indirect) effects of the policy. For this purpose, we also report matching-DID estimates computed on the common support (see Villa, 2016). These estimates can be roughly interpreted as the behavioral effects estimated by matching the post-policy students with those from the pre-policy period. Also in this case, as for our baseline estimates, these do not account for the "cream-skimming" effect, which however might have been an (intended or unintended) effect of the reform. Kernel matching is performed on all covariates except for the income brackets (that cannot be balanced between DSU and non-DSU students since student aid eligibility is only based on income in the first year). The DID model includes the same covariates as the baseline model. Table 6 reports the matching-DID estimates.

The matching-DID estimates are very close to our baseline estimates. We find positive and significant effects on first-year credits earned and credits earned by August 10 and November 30 as well as on the probability of on-time graduation and with at most one year of delay. In contrast to our baseline DID estimates, we find a statistically significant negative effect of 2.7 pp (about -2.8%) on the probability of enrolling in the second year and statistically insignificant effects on GPA. As discussed in the next Section, the negative effect on second-year enrollment is driven by lower-ability students. In general, the matching-DID estimates have magnitudes quite comparable to our baseline estimates, suggesting again that changes in observable characteristics are unlikely to have an important impact on the DID estimates.

### 6.3. *Effect heterogeneity: Winners and losers from tightening the merit requirements*

In this Section, we explore whether the effects of the reform are related to the different initial ability of students, i.e. if the policy had heterogeneous effects on the distribution of student ability. To this end, we divide students into three groups using the quartiles of

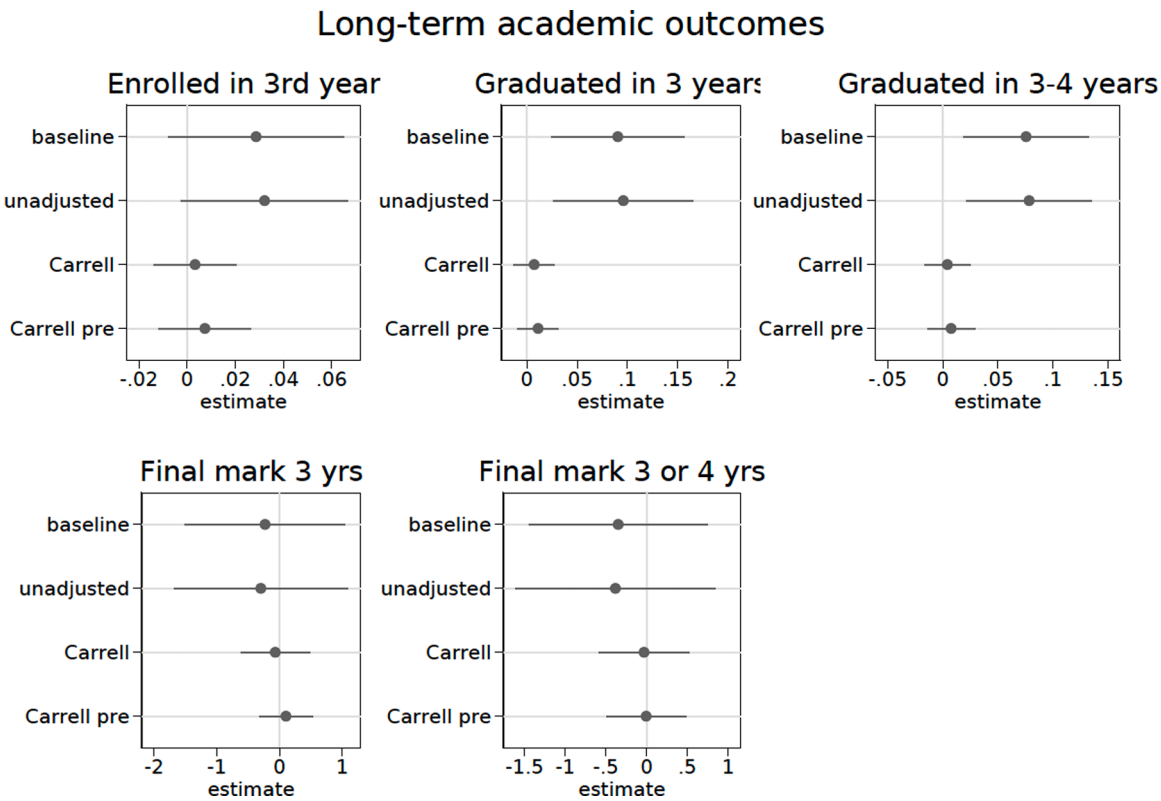


**Fig. 2.** The effect of time-varying covariates – Long-term outcomes. *Note.* Each graph reports for each dependent variable the coefficient on the $DSU_i$ $*After_t$ interaction estimated on the baseline model ("baseline") controlling for the variables in Table 2, the unadjusted model (i.e. omitting the covariates, "unadjusted") and using the predictions of the student outcomes based on the covariates estimated on the whole period ("Carrell") or only in the pre-policy period ("Carrell pre"). The 2010 cohort is excluded. For all models confidence intervals are computed using White robust standard errors.

the admission test score distribution (by entry-cohort and field): "lower ability" are those falling in the first quartile, "medium ability" are those in the second and third quartile, and "higher ability" are those in the fourth quartile.[21]

The results for short-term outcomes are reported in Table 7. The most striking finding is that lower-ability students were negatively affected by the reform. Although the reform does not seem to have had any significant effect on the number of credits earned, it lowered GPAs at the different dates by about half grade (an effect between -2.3% and -2.5% at baseline) and more importantly induced a 10-pp decrease (a 10% reduction at the baseline) in the probability of enrolling in the second year. In the Italian system, students can sit an exam in several exam sessions during the year and can refuse grades and retake the exam if they are not satisfied with their performance. The higher merit requirements may have influenced students, especially those of lower ability, into taking exams by August 10th and accepting low grades. As for the number of credits earned, the results also suggest that the reform had the largest effects on higher- and medium-ability students. For instance, credits earned by August 10th increased on average by 3.5 (7.2%) and 4 (8.9%) for higher- and medium-ability students in aid, respectively. In other words, the average positive effects shown in Table 2 were driven by medium- and higher-ability students. All in all, our results of the inequality-increasing effects of merit requirements are in line with Scott-Clayton and Schudde (2020), although our study specifically focuses on the effect of increasing these requirements and not of failing them.

Table 8 shows the results for long-term outcomes. Higher- and medium-ability students experienced gains of about 11 pp (corresponding to an increase of 15% and 14% for higher- and medium ability students, respectively) in the probability of graduating with

[21] We explicitly allow for the treatment effects to be heterogeneous by including the interactions with quartile of ability in the model. By doing this, we are looking at the effects of the policy conditional on being in each ability quartile; thus, DSU and non-DSU students in a given quartile before the reform are compared with those in the same quartile after the reform. Changes in ability quartiles among DSU students would be captured by quartile of ability indicators and by the ability score that is included linearly in the model.

**Table 6**
Matching-DID estimates

| Short-term outcomes | Coeff. | S.E. | N. | Long-term outcomes | Coeff. | S.E. | N. |
|---|---|---|---|---|---|---|---|
| Credits earned 1st year | 1.436* | (0.831) | 26,414 | Enrolled in the 3rd year | 0.028 | (0.019) | 26,414 |
| Credits earned Aug 10 | 2.654*** | (0.862) | 26,414 | Graduated in 3 yrs | 0.087** | (0.034) | 26,414 |
| Credits earned Nov 30 | 1.869** | (0.828) | 26,414 | Graduated in 3 or 4 yrs | 0.073** | (0.029) | 26,414 |
| GPA 1st year | 0.234 | (0.169) | 23,477 | Final mark 3 yrs | 0.036 | (0.673) | 8,588 |
| GPA Aug 10 | 0.264 | (0.171) | 23,252 | Final mark 3 or 4 yrs | -0.456 | (0.574) | 15,599 |
| GPA Nov 30 | 0.238 | (0.169) | 23,479 | | | | |
| Enrolled in the 2nd year | -0.027** | (0.013) | 26,414 | | | | |
| Control variables | Yes | | | | Yes | | |

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses.
*Note.* The Table shows the matching-DID estimates of the $DSU_i$*$After_t$ coefficient of equation (1) on the common support. Post-reform cohorts (After) are those enrolled after 2010/2011. The number of observations varies across student outcomes as GPA is only available for exams that attribute a grade ("pass" or "fail" grades are excluded). Among the long-term outcomes, graduation marks are conditional on graduation. The 2010 cohort is excluded. Control variables include gender, age, entry test score, income brackets, residence and college major. Kernel matching is carried out on the same variables, except income brackets. N. is the number of observations in the common support.

at most one year of delay. In contrast, no gain is observed for lower-ability students. For this group, the reform, if anything, worsened final graduation marks, especially for those graduating within the normative time to completion (by about -4 points; a decrease of 4%). This effect is consistent with our findings on the first-year GPA for lower-ability students. It is likely that fearing to lose financial aid, these students reduced their "reservation grades" and were more willing to accept lower grades compared to the pre-policy period.[22,23] As discussed in Section 2, lower grades may, in turn, affect graduates' labor market outcomes and for this reason some students might have decided to drop out from PoliMi after their first year of enrollment.[24]

## 7. Discussion and concluding remarks

There has been a proliferation of need-based aid programs attaching merit-based requirements to financial aid as incentives to improve student success (Anderson et al., 2020). While the need-based component of these programs helps more disadvantaged students to get access to higher education, the goal of merit-based requirements is to increase the efficiency of student aid by raising student effort and performance. Excessively high merit-based requirements, though, may induce some students to fail the standards set for aid renewal or for not having to repay the aid, and drop out from the university. In other words, merit-based requirements may create an "equity-efficiency trade-off." Despite the importance of this issue, research studying the effects of increasing merit-based requirements in need-based aid is still scant.

In this paper, we aim to fill this gap using quasi-experimental evidence provided by a reform that increased first-year credit requirements for student aid renewal by 40% in the most populous Italian region (Lombardy). Our analysis on administrative data from an Italian university mainly offering STEM degrees (*Politecnico di Milano*) shows that the reform increased student performance along several dimensions. Our DID estimates show average positive effects of the reform both on first-year student outcomes, namely on the number of credits earned, but also in the longer term, with an increase of on-time graduation. Thus, some students are likely to have benefited also economically from the reform, as a shorter time-to-degree is generally associated with better employment and earnings outcomes (see, for instance, Aina and Casalone 2020 in the Italian context).[25]

By contrast, our heterogeneity analysis by student ability (measured by the admission test score) uncovers winners and losers from the reform. Indeed, the reform decreased the probability of lower-ability students (i.e. those at the bottom quartile of the admission test score) re-enrolling in the second year by about 10 pp. We put forward two main hypotheses for explaining this finding. First, our results suggest that completely losing financial support may have decreased college persistence of lower-ability students who did not meet the

[22] It is worth noting that when measuring the probability of graduation, the outcome takes on a value of zero both for dropouts and delayers, i.e. those who are still enrolled. This is the reason why one could find a significant negative effect on the probability of enrolling in the second year and still not find any significant negative effect on the probability of on-time graduation, i.e. the increase in drop-out (entailing a lower number of graduates on-time) could be compensated by an increase of students who persist and graduate on-time.

[23] We also investigated gender differences in the effect of the reform. Results are reported in the online Appendix. We explored heterogeneous effects by gender including interaction terms. We find that the reform had similar effects for males and females on the number of credits earned by August 10$^{th}$ (online Appendix Table A1). As for the long-term outcomes (online Appendix Table A2), the reform increased the probability of on-time graduation for both genders in a similar fashion, while the probability of graduating with at most one-year delay only increased for men (about 10 pp).

[24] Our estimates may suffer from small cell size problems. The number of lower-ability DSU students is 74 and 51 after and before the reform, respectively. The number of medium-ability DSU students is 265 after and 162 before the reform, and that of high-ability DSU students is 147 after and 162 before the reform.

[25] Although assessing causality in this context is difficult, longer graduation times may signal lower graduates' skills to employers (Aina and Pastore, 2020).

merit requirements. Second, the reform also reduced their GPA as they had to accept lower grades to fulfill the merit requirements (in the Italian HE system students can retake the exams if they are not satisfied with the grades received). Given that lower grades would have affected the final graduation mark, lower-ability students might have left PoliMi to enroll in a less selective HE institution.

A limitation of single-institution data is that we cannot distinguish between dropouts from HE and dropouts from PoliMi who enroll in another HE institution (see, for instance, Carrieri et al., 2015). In the best-case scenario, the reform is likely to have worsened the academic performance in the student careers of lower-ability students or forced them to leave a selective institution, raising equity concerns. In the worst-case scenario, the policy may have induced lower-ability students to drop out from higher education, with even greater individual and societal economic losses. We attempt to understand the costs of dropping out of PoliMi by using aggregate information of students who transfer to another Italian HE institution within 12 or 15 months relative to first enrollment. The online Appendix Table A6 shows that those who transfer within 12 months relative to first enrollment represent between 3.7% and 7.7% of all students who drop out of PoliMi.[26] Among those who transfer to another Italian HE institution, between 30% and 50% remain in STEM majors, although only 1% to 4% enroll in another HE institution that is as prestigious as PoliMi.[27] This suggests that those who leave PoliMi potentially worsen their academic and labor market returns by enrolling in a less prestigious college even if they persist in STEM majors and increasing their time-to-graduation (if they cannot transfer credits).

Although lowering participation of lower-ability low-income students might be the intended goal of tightening merit requirements, there are downsides. First, this may have equity implications since students with similar ability levels from more advantaged backgrounds were not subjected to the same performance requirements. It is therefore key to not set the bar of "merit"—which only applies to disadvantaged students, namely those receiving financial aid—too high. Second, our measure of student ability is based on performance in a university admission test. This may depend not only on innate ability or "merit" (e.g. student effort) but also on inputs that are likely to be affected by socio-economic backgrounds, such as high school quality or out-of-school inputs, e.g. private tutoring classes (Sianou-Kyrgiou, 2008; Guimarães and Sampaio, 2013). Thus, our results may be interpreted also as stricter merit requirements penalizing more students who were disadvantaged from the start, i.e. when entering HE. Finally, as we argue below, finding negative effects for disadvantaged students enrolled in highly selective institutions (in which the student intake is likely to be positively selected) and for STEM degrees (that offer high labor market returns) may be particularly worrying from an equity perspective. In other words, too high merit requirements may push disadvantaged students out of the most prestigious HE institutions and from STEM degrees.

Our paper shares the same limitations as other studies that rely on analysis of single institutions (e.g., Montalbán, 2019; Leuven et al., 2010) or single states (e.g. Scott-Clayton, 2011; Scott-Clayton and Schudde, 2020) to investigate the effect of financial aid on student performance: these effects are likely to depend on how financial aid is designed and the institutional features of HE systems. Merit requirements are designed differently across countries. In our study, these requirements were defined only in terms of credits to be earned, corresponding to about 40% of annual credits before the reform and 60% after the reform. These thresholds are lower than those of Becas de Carácter General that were between 60% and 80% of annual credits depending on the college major and were raised to between 85% and 100% (or between 65% to 90% when combined with GPA requirements) after the reform studied by Montalbán (2019). Satisfactory academic progress for the Pell Grant requires students to complete two thirds of the credits they attempt (i.e. about 67%) and a GPA of 2.0 or higher (Schudde and Scott-Clayton, 2016). Thus, student aid systems are very heterogeneous, not only in terms of the number of credit requirements but also in the inclusion of minimum GPA requirements. Moreover, the percentage of beneficiaries and the amount of the grants also vary across European countries (European Commission, 2015). In this context, it is difficult for single-institution and single-country studies to have high external validity. Yet, this paper is one of the first to estimate the effect of increasing merit requirements. Even though our study refers to a single institution (PoliMi), our findings are potentially applicable to similar policies introduced in other large, selective, and STEM-oriented Higher Education institutions.

PoliMi is located in a major town in Northern Italy (Milan), where students face very high costs of living. Thus, losing financial aid may be particularly harmful and some students may either drop out from HE or enroll in universities in other cities and regions to continue their tertiary education. This could make our results for lower-ability students an "upper-bound" estimate (in absolute value). Moreover, PoliMi is a highly selective institution offering STEM degrees that yield higher returns in the labor market and attract highly motivated and higher achieving students, relative to the average Italian university. This probably makes the positive effects on performance for higher- and medium-ability students an "upper-bound", but a "lower bound" for the probability of not enrolling in the second year for lower-ability students. Indeed, given the perceived high returns to attending a more selective institution, lower-ability students may be more likely to persist in college even after losing student aid compared to those enrolled in less selective institutions, for which the higher costs induced by similar reforms might greatly exceed the returns of persisting in college. Some specific features of the Italian HE system might also influence our findings. The reform was partially implemented to encourage students to sit exams and accept grades as early as possible to meet the credit requirements by August, to reduce time-to-degree. Relative to other HE systems, Italian students have the academic advantage of being allowed to retake the same exam several times during the academic year which might help them meet the merit requirements. Therefore, the inequality induced by increasing the merit requirements may be larger in other countries.

[26] The percentage of students who transfer out to PoliMi within 12 months since first enrollment halves after 2010, the year in which the number of DSU students also declined. Thus, one could put forward that DSU students are relatively less likely to transfer compared to non-DSU students. However, without individual level data on student transfers we cannot further explore this hypothesis.

[27] PoliMi consistently ranks among the top-paying universities in the Italian labor market (Ciani and Mariani, 2014), both in general and among HE institutions offering STEM degrees.

**Table 7**
Heterogeneous effects by student ability (i.e. admission test score) on short-term outcomes

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Variables | CFU first year | CFUs Aug 10 (1st year) | CFUs Nov 30 (1st year) | GPA 1st year | GPA Aug 10 (1st year) | GPA Nov 30 (1st year) | Enrolled in the 2nd year |
| DSU | 6.693 | 3.959 | 5.533 | 0.278 | 0.025 | 0.257 | -0.565*** |
| | (7.971) | (6.934) | (7.923) | (0.705) | (0.753) | (0.705) | (0.038) |
| DSU * higher ability | -8.117*** | -6.944*** | -7.712*** | -0.179 | -0.213 | -0.153 | -0.207*** |
| | (1.644) | (1.891) | (1.631) | (0.284) | (0.299) | (0.284) | (0.018) |
| DSU * medium ability | -3.935*** | -3.665** | -3.785*** | -0.634*** | -0.577** | -0.628*** | -0.127*** |
| | (1.398) | (1.649) | (1.400) | (0.243) | (0.255) | (0.243) | (0.015) |
| DSU * After (a) | 1.498 | 1.913 | 2.153 | -0.556* | -0.610* | -0.546 | -0.104*** |
| | (2.313) | (2.438) | (2.315) | (0.336) | (0.340) | (0.336) | (0.039) |
| DSU * higher ability * After (b) | 0.102 | 1.532 | -0.413 | 0.914** | 1.145** | 0.894** | 0.098** |
| | (2.624) | (2.796) | (2.623) | (0.444) | (0.448) | (0.444) | (0.042) |
| DSU * medium ability * After (c) | 1.433 | 2.108 | 1.493 | 0.962** | 0.942** | 0.971** | 0.114*** |
| | (2.527) | (2.639) | (2.526) | (0.406) | (0.411) | (0.406) | (0.043) |
| Control variables | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N. | 26,689 | 26,689 | 26,689 | 23,583 | 23,346 | 23,585 | 26,689 |
| (a) + (b) [higher ability] | 1.600 | 3.446** | 1.740 | 0.358 | 0.534** | 0.348 | -0.006 |
| *p*-value | 0.232 | 0.016 | 0.190 | 0.220 | 0.070 | 0.234 | 0.733 |
| (a) + (c) [medium ability] | 2.931*** | 4.021*** | 3.646*** | 0.407** | 0.332 | 0.425** | 0.011 |
| *p*-value | 0.007 | 0.000 | 0.001 | 0.080 | 0.162 | 0.067 | 0.562 |

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses.
*Note.* DSU students are students receiving aid. Post-reform cohorts (After) are those enrolled after 2010/2011. The number of observations varies across student outcomes as GPA is only available for exams that attribute a grade ("pass" or "fail" grades are excluded). Among the long-term outcomes, graduation marks are conditional on graduation. Control variables include gender, age, entry test score, income brackets, residence and college major. The 2010 cohort is excluded. Lower-ability, medium-ability and higher-ability students are those belonging to the 1st, 2nd -3rd and 4th quartiles of the entry-test score distribution, respectively. While the (DSU * After) interaction captures the effect of the reform for lower-ability students, effects for medium- and higher-ability students are reported in the bottom part of the table. N. is the number of observations.

**Table 8**
Heterogeneous effects by student ability (i.e. entry test score) on long-term outcomes

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Variables | Enrolled in the 3rd year | Graduated in 3 yrs | Graduated in 3/4 yrs | Final mark 3 yrs | Final mark 3/4 yrs |
| DSU | 0.176 | -0.332*** | -0.129 | 2.193** | 0.507 |
| | (0.171) | (0.088) | (0.161) | (0.977) | (1.277) |
| DSU * higher ability | -0.197*** | 0.021 | -0.211*** | -0.603 | 0.618 |
| | (0.033) | (0.064) | (0.060) | (1.152) | (1.019) |
| DSU * medium ability | -0.157*** | 0.007 | -0.121** | -1.239 | -0.806 |
| | (0.032) | (0.058) | (0.056) | (1.145) | (0.953) |
| DSU * After (a) | -0.071 | 0.032 | -0.063 | -3.904** | -2.550* |
| | (0.052) | (0.080) | (0.082) | (1.954) | (1.416) |
| DSU * higher ability * After (b) | 0.084 | 0.048 | 0.175* | 3.252 | 1.339 |
| | (0.057) | (0.098) | (0.093) | (2.162) | (1.672) |
| DSU * medium ability * After (c) | 0.154*** | 0.083 | 0.169* | 4.718** | 3.245** |
| | (0.058) | (0.093) | (0.091) | (2.167) | (1.612) |
| Control variables | Yes | Yes | Yes | Yes | Yes |
| N. | 26,689 | 26,689 | 26,689 | 8,829 | 15,636 |
| (a) + (b) [higher ability] | 0.013 | 0.080 | 0.112** | -0.652 | -1.211 |
| *p*-value | 0.611 | 0.155 | 0.011 | 0.489 | 0.180 |
| (a) + (c) [medium ability] | 0.082*** | 0.115** | 0.106** | 0.814 | 0.695 |
| *p*-value | 0.002 | 0.019 | 0.010 | 0.395 | 0.383 |

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses.
*Note.* DSU students are students receiving aid. Post-reform cohorts (After) are those enrolled after 2010/2011. The number of observations varies across student outcomes as GPA is only available for exams that attribute a grade ("pass" or "fail" grades are excluded). Among the long-term outcomes, graduation marks are conditional on graduation. Control variables include gender, age, entry test score, income brackets, residence and college major. The 2010 cohort is excluded. Lower-ability, medium-ability and higher-ability students are those belonging to the 1st, 2nd -3rd and 4th quartiles of the entry-test score distribution, respectively. While the (DSU * After) interaction captures the effect of the reform for lower-ability students, effects for medium- and higher-ability students are reported in the bottom part of the table. N. is the number of observations.

Our findings suggest that introducing stronger merit requirements may discourage students from more disadvantaged family backgrounds from persisting in college. This is particularly worrying for highly selective institutions in which the share of low-income students is already quite low. Our analysis shows that although these reforms appear to change the composition of students by increasing the average ability of lower-income students, the self-selection effect is not sufficiently large to deter students from dropping out of selective HE institutions. Although effect sizes are difficult to compare across different contexts and institutional settings, it is

worth noting that our findings on the redistributive effects of *increasing merit requirements* are qualitatively consistent with the estimated effects of failing SAP by Schudde and Scott-Clayton (2016) and Scott-Clayton and Schudde (2020).

A final important conclusion of our study is that as student aid programs are generally designed as "all or nothing" packages when it comes to meeting merit requirements, policy makers should consider second chances for students who lose funding because they do not meet these requirements the first time around. As suggested by Jones et al. (2022), a redesign of merit-based renewal requirements under the form of a sliding scale, in which students only partially lose financial aid depending on their performance may more effectively increase aid efficiency without exacerbating inequality.

## Acknowledgements

We gratefully acknowledge comments and suggestions from the Editor Prof. Robert M. Sauer, from three anonymous reviewers, and from participants in seminars and conferences in which earlier versions of this paper were presented, among which the 2019 Association for Education Finance and Policy (AEFP) Conference (Kansas City), IRVAPP (Trento), the 2019 annual conference of the Italian Association of Labour Economists (AIEL), the 2019 workshop "The Economics of Higher Education" ZEW (Mannheim), the Ninth Italian Congress of Econometrics and Empirical Economics (2021) and the 2021annual meeting of the Italian Economics Association (SIE). The usual disclaimer applies. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.euroecorev.2022.104164.

## References

Aina, C., Baici, E., Casalone, G., 2011. Time to degree: Students' abilities, university characteristics or something else? Evidence from Italy. Education Economics 19 (3), 311–325.

Aina, C., Casalone, G., 2020. Early labor market outcomes of university graduates: Does time to degree matter? Socio-Economic Planning Sciences 71 (C), 100822.

Aina, C., Pastore, F., 2020. Delayed graduation and overeducation in Italy: A test of the human capital model versus the screening hypothesis. Social Indicators Research 152 (2), 533–553.

Altonji, J.G., Kahn, L.B., Speer, J.D., 2014. Trends in earnings differentials across college majors and the changing task composition of jobs. American Economic Review 104 (5), 387–393.

Anderson, D.M., Broton, K.M., Goldrick-Rab, S., Kelchen, R., 2020. Experimental evidence on the impacts of need-based financial aid: Longitudinal assessment of the Wisconsin Scholars Grant. Journal of Policy Analysis and Management 39 (3), 720–739.

Angrist, J., Oreopoulos, P., Williams, T., 2014. When opportunity knocks, who answers? New evidence on college achievement awards. Journal of Human Resources 49 (3), 572–610.

Argentin, G., 2010. University pathways and graduate labour market outcomes in Italy: What matters where? Italian Journal of Sociology of Education 2 (2), 107–147.

Belot, M., Canton, E., Webbink, D., 2007. Does reducing student support affect scholastic performance? Evidence from a Dutch reform. Empirical Economics 32 (2-3), 261–275.

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

Carrieri, V., D'Amato, M., Zotti, R., 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.

Ciani, E., Mariani, V., 2014. How the labour market evaluates Italian universities. Bank of Italy Occasional Paper. Rome: Bank of Italy.

Civera, A., Cattaneo, M., Meoli, M., Paleari, S., Seeber, M., 2020. Universities' responses to crises: The influence of competition and reputation on tuition fees. Higher Education. https://doi.org/10.1007/s10734-020-00622-2.

Dynarski, S., Scott-Clayton, J., 2013. Financial aid policy: Lessons from research, 23. The Future of Children, pp. 67–91.

European Commission/EACEA/Eurydice, 2015. *National Student Fee and Support Systems in European Higher Education –2014/15. Eurydice* – Facts and Figures. Luxembourg: Publications Office of the European Union.

Almalaurea Consortium, 2021. XXIII Indagine (2021) - Profilo dei Laureati 2020 (Graduates' profile 2020). Available online https://www.almalaurea.it/universita/profilo/profilo2020.

EUROSTUDENT (2018). Le condizioni di vita e di studio degli studenti universitariu 2016-2018 (Life and study conditions of students, 2016-2018), accessed online http://www.eurostudent.it/i-rapporti-dellindagine-italiana/.

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

Ghignoni, E., 2017. Family background and university dropouts during the crisis, 73. The case of Italy. Higher Education, pp. 127–151.

Guimarães, J., Sampaio, B., 2013. Family background and students' achievement on a university entrance exam in Brazil. Education Economics 21 (1), 38–59.

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

Leuven, E., Oosterbeek, H., van der Klaauw, B., 2010. The effect of financial rewards on students' achievement: Evidence from a randomized experiment. Journal of the European Economic Association 8 (6), 1243–1265.

Lewis, J.B., Linzer, D.A., 2005. Estimating regression models in which the dependent variable is based on estimates. Political analysis 13 (4), 345–364.

Mealli, F., Rampichini, C., 2012. Evaluating the effects of university grants by using regression discontinuity designs. Journal of the Royal Statistical Society: Series A (Statistics in Society) 175 (3), 775–798.

Modena, F., Rettore, E., Tanzi, G.M., 2020. The effect of grants on university dropout rates: Evidence from the Italian case. Journal of Human Capital 14 (3), 343–370.

Montalbán, J., 2019. Countering moral hazard in higher education: The role of performance incentives in need-based grants. ffhalshs-02160365v2f. (conditionally accepted at the Economic Journal).

Piazzalunga, D., 2018. The gender wage gap among college graduates in Italy. Italian Economic Journal 4 (1), 33–90.

Schudde, L., Scott-Clayton, J., 2016. Pell Grants as performance-based scholarships? An examination of satisfactory academic progress requirements in the nation's largest need-based aid program. Research in Higher Education 57 (8), 943–967.

Scott-Clayton, J., Schudde, L., 2020. The consequences of performance standards in need-based aid: Evidence from Community Colleges. Journal of Human Resources 55 (4), 1105–1136.
Scott-Clayton, J., 2011. On money and motivation. A quasi-experimental analysis of financial incentives for college achievement. Journal of Human Resources 46 (3), 614–646.
Sianou-Kyrgiou, E., 2008. Social class and access to higher education in Greece: Supportive preparation lessons and success in national exams. International Studies in Sociology of Education 18 (3-4), 173–183.
Zeldow, B., Hatfield, L.A., 2021. Confounding and regression adjustment in difference-in-differences studies. Health services research 56 (5), 932–941.