I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15456

COVID-19 and the Gender Gap in University Student Performance

Massimiliano Bratti
Enrico Lippo

JULY 2022

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15456

COVID-19 and the Gender Gap in University Student Performance

Massimiliano Bratti
University of Milan, IZA and LdA

Enrico Lippo
University of Milan

JULY 2022

ISSN: 2365-9793

IZA – Institute of Labor Economics

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

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

www.iza.org

ABSTRACT

COVID-19 and the Gender Gap in University Student Performance*

The gendered impact of the COVID-19 pandemic has been observed in many domains, such as labor market outcomes and mental health. One sector that was particularly disrupted by the pandemic was education, owing to the need to close educational institutions and move all learning activities online. In this paper, we investigate the gender gap in university student performance, focusing on a large public university located in one of the European regions most affected by the first pandemic wave (Lombardy, in Northern Italy). Despite concerns that the pandemic might have had a heavier toll on the educational performance of female students, our empirical analysis shows that the gender gap in student progression (number of credits earned) was not affected by the pandemic and that in some college majors (social sciences and humanities) women even improved their GPA relative to men.

JEL Classification: I21, I23

Keywords: COVID-19, university, student performance, gender gap

Corresponding author:
Massimiliano Bratti
Dipartimento di Economia
Università degli studi di Milano
Via Conservatorio, 7
20122 Milano
Italy
E-mail: massimiliano.bratti@unimi.it

* The authors are deeply indebted to several offices and colleagues of the University of Milan that supported this project in several forms, mainly by providing data and useful information. In particular Matteo Turri, Yuri Pasquinelli, Marialuisa De Francesco, Silvia Spazzacampagna, Laura Carradori, Stefania Gavina, Guido Chiantoni, Daniela Bagnati and Diego Domenis. Without their collaboration, this project would not have been possible. Comments received by several colleagues are gratefully acknowledged. The usual disclaimers apply.

1 Introduction

A new word was forged to indicate the gendered effects of the COVID-19 pandemic on economic activity: "shecession", a recession that hits women more heavily than men. In fact, there is plenty of evidence that the toll of COVID-19 was higher for women, not only in the economic sphere, such as for employment (Alon et al., 2020; Graeber et al., 2021; and, for Italy, Del Boca et al., 2020) but also for mental health (Etheridge and Spantig, 2020; Zamarro and Prados, 2021).

Although scholars have investigated the learning losses associated with, or caused by, the COVID-19 pandemic both at school (Kuhfeld et al., 2020; Engzell et al., 2021) and at the university level (e.g., Orlov et al., 2021; De Paola et al., 2022) and have highlighted differences by student socio-economic status (Aucejo et al., 2020; Rodríguez-Planas, 2022), less is known about the evolution of gender gaps in education during the pandemic.[1] It must be kept in mind that as far as university performance is concerned, women are not necessarily suffering a negative gap compared to men. On the contrary, women's academic performance is oftentimes above that of men. Thus, a larger negative impact of COVID-19 on the performance of female students might entail a reduction in women's advantage, and of the gender gap.

Several mechanisms may lead to an uneven effect of the pandemic on student performance by gender. On the one hand, previous research has shown that the mental health of women was severely affected (Prowse et al., 2021), causing a relative improvement in male students' performance. School closures may have increased the workload at home required of female students who have children and family responsibilities (Del Boca et al., 2020; Sevilla and Smith, 2020; Giurge et al., 2021; Zamarro and Prados, 2021) or who have young siblings, reducing the time available for education.[2] COVID-19 marked the beginning of a new way of teaching that is heavily reliant on ICT. Given the well-known gender gap in ICT aptitudes and use, women might be less involved in on-line learning compared to men (Meelissen and Drent, 2008). Yet not all effects of COVID-19 are likely to have improved the relative performance of men compared to women. Indeed, remote learning

[1] For the purpose of the current paper, gender refers to an individual's biological sex as recorded in university administrative data.

[2] However, given the typical age of students enrolled in higher education, we do not expect this channel to be prevalent.

requires a substantial amount of self-organization and self-discipline (i.e., "self-regulated learning") by students, characteristics that are more common among women (Liu et al., 2021). Moreover, although the pandemic is likely to have reduced peer-group effects (Agostinelli et al., 2022), it might also have removed some of the negative effects of excess peer competition on learning. Delaney and Devereux (2021), for instance, show that in mixed-sex schools, students' within-cohort rank in particular subjects is more important for men than for women in choosing to study STEM, while this is not the case for same-sex schools. They offer as a possible explanation the fact that men's behavior may be affected by the presence of students of the opposite sex; for instance, they may behave more competitively or display more masculine traits when surrounded by women.

Although it was not their main focus, a few papers have investigated the heterogeneous effects of the pandemic on tertiary students' performance by gender. De Paola et al. (2022) used student administrative data from a university in Southern Italy. They found a negative effect of the pandemic on *credits earned* but no significant differences across gender. Bonaccolto-Töpfer and Castagnetti (2021) investigate the effect of COVID-19 on the *average grades of students* (GPA) using administrative data on a university in Northern Italy. In bachelor degrees, the estimated effects are nil when considering all exams and positive but small when considering only compulsory exams, but their results show statistically significant small negative effects when considering multiple pre-treatment periods. In all cases, the effects are very similar across genders. In a comparative study of three universities in Italy, Turkey, and Sweden (Casalone et al., 2021), the authors exploit different mitigation policies in the three countries and show that although the pandemic reduced the *probability of passing exams*, the effect was partially mitigated by lockdown policies. When the analysis is split by gender, the effects were positive for females and non-significant for males. The authors suggest an increase in the time available for studying as a possible explanation.

Orlov et al. (2021) use data from intermediate economics courses taught at four US institutions. Albeit the pandemic negatively impacted student test scores, the authors found no significant difference across gender. Using a survey of about 1,500 students from Arizona State University, Aucejo et al. (2020) study the impact of COVID-19 on a number of subjective academic and labor market outcomes. Students were generally negatively impacted, but differences across gender emerge only for a limited number of outcomes, namely semester GPA and the probability of changing major, which reduced and increased more for men, respectively, and the willingness to take online classes in the future, which was negatively affected by COVID-19 for men only.

In this paper, we contribute to this scant literature and investigate how gender differences in university student progression evolved during the first wave of the pandemic. Compared to most extant work, which uses survey data with a relatively low number of respondents or with potential representativeness issues (e.g., Aucejo et al., 2020) or administrative data from single degrees or courses (e.g., Kofoed et al., 2021; Orlov et al., 2021; Supriya et al., 2021), we use administrative data on the whole student population of a large public university in Northern Italy, the University of Milan (UniMI hereafter).

UniMI represents an interesting case study for a number of reasons. First, it is one of the largest universities in Italy, and we are thus able to use large samples in the estimation, offering a good level of statistical power.[3] Second, UniMI offers a wide menu of degrees spanning several college majors. UniMI supplies degrees in STEM, humanities, law, social sciences, and medicine and dentistry. Thus, we can provide evidence for certain college majors that are typically male-dominated (e.g., STEM) and others that are female-dominated (e.g., humanities) in terms of the percentage of male and female students enrolled. This large supply of degrees also entails high variability in student characteristics and degree entry requirements, potentially increasing the external validity of our results.

Finally, the region of Lombardy, where UniMI is located, was among the European regions most heavily affected by the first wave of the COVID-19 pandemic. This must be kept in mind because compared to other studies, which tend to mainly estimate the effect of switching to online teaching (such as De Paola et al., 2022, which was based in Calabria where, as in most of Southern Italy, the health emergency was much less severe during the first wave of the pandemic compared to the northern parts of the country), for UniMI the estimated effects are likely to capture the impact of both online teaching and other impacts of the pandemic. Thus, albeit not exempt from external

[3] UniMI counted 61,279 students enrolled in the 2020/21 academic year, ranking fourth among Northern Italy's higher education institutions in terms of the number of students after the University of Bologna (81,931 students), the University of Turin (77,809 students), and the University of Padua (63,061 students); Source: University and Research open data).

validity concerns (being related to one higher education institution only), our results have higher external validity than many other studies in the literature, which are based on small samples or a single/few university courses. Other studies have used administrative data on the entire student population from single institutions, but without a specific focus on gender gaps (e.g., Bonaccolto-Töpfer and Castagnetti, 2021; De Paola et al., 2022; Rodríguez-Planas, 2022).

Administrative data have some advantages. Namely, we can rely on the whole student population, in contrast to studies using survey data, and student administrative records provide precise and objective (i.e., not self-reported) measures of student academic performance. As a downside, administrative sources — mainly providing information on student performance and student socio-demographic characteristics — are unsuitable for an in-depth analysis of the potential mechanisms explaining the evolution of gender differences after the pandemic, which would require detailed information on student behavior. For this reason, this paper only focuses on the analysis of gender differences in university student performance before vs. after the first COVID-19 wave, providing an effect heterogeneity analysis by college major but without investigating the potential mechanisms.[4]

The remainder of this paper proceeds as follows. Section 2 introduces the empirical strategy. Section 3 presents the institutional setting and the data used in the analysis. Section 4 discusses the main results. Finally, Section 5 draws conclusions.

2 Empirical strategy

We parametrically implement a difference-in-difference-in-differences (or triple-difference, DDD hereafter) research design (Gruber, 1994; Olden and Møen, 2020). Our estimated model reads as follows:

[4] De Paola et al. (2022) using matched administrative-survey data show that the negative impact of the pandemic on student performance was larger on students showing present-bias.

$$y_{ijt} = \alpha_0 + \alpha_1 sem2_t + \alpha_2(sem2_t \times a.y.\,2019/20_j) + \alpha_3(sem2_t \times female_i) + \alpha_4(female_i \times a.y.\,2019/20_j) + \alpha_5(sem2_t \times female_i \times a.y.\,2019/20_j) + a.y._j + u_i + \epsilon_{ijt}, \tag{1}$$

where y_{ijt} is the number of credits earned for student i in period t of academic year j. Each period t is the combination of the semester (first or second) and the academic year j ($t = 1, ..., 6$).[5] The model includes, on the right-hand-side, a second-semester indicator $sem2_t$ (each academic year is split into two semesters);[6] an interaction between the second semester indicator and the $a.y.\,2019/20_j$ indicator, which is one for the academic year affected by COVID-19 (2019-2020) and zero otherwise; double interactions between the female indicator, $female_i$, and the second semester ($sem2_t$) and academic year 2019-2020 ($a.y.2019/20_j$) indicators, respectively; the triple interaction $sem2_t \times female_i \times a.y.\,2019/20_j$, whose coefficient α_5 captures the triple difference between the second and first semesters, men and women, and the pre- and post-COVID-19 period; academic year fixed effects ($a.y._j$). u_i are student fixed effects, and ϵ_{ijt} is an idiosyncratic error term. We allow for the error terms of observations to be correlated within degree courses, clustering the estimates at the degree-course level.[7] α_5 is our parameter of interest and captures the differential impact of the pandemic on women *vis-à-vis* men. We estimate variants of equation (1) in which the number of credits is replaced with students' semester GPA.

Our research design is a mix of the one followed by Rodríguez-Planas (2022), which investigates socio-economic gaps in the effect of the pandemic on university student performance, and of De Paola et al. (2022), which investigates the impact of COVID-19 on performance in an Italian

[5] Since we start counting from the academic year 2017 (see Section 3), $t = 1$ for the first semester of 2017, $t = 2$ for the second semester of 2017, $t = 3$ for the first semester of 2018, and so on.

[6] So $sem2_t = 1$ in periods $t = 2, 4, 6$ and $sem2_t = 0$ in periods $t = 1, 3, 5$. This is the case for all UniMI schools except for the School of Social, Political and Economic Sciences (SPES), for which the academic year is split into three quarters. In this case, data from the second and third quarters are pooled to build an aggregate comparable to semesters. More precisely, the first quarter corresponds to the first semester and the second and third quarters to the second semester. This split is motivated by the fact that at SPES the second and third quarters of the academic year 2019-2020 were affected by COVID-19.

[7] This allows not only an individual's observations to be correlated over time, but also observations related to different students to be correlated within each degree course. This clustering is more conservative than that at the individual level, which excludes correlation across individuals enrolled in the same degree.

university.[8] Like Rodríguez-Planas (2022) and De Paola et al. (2022), we focus on an unbalanced panel of students (see Section 3 for details). Consistent with Rodríguez-Planas (2022), we include student fixed effects in our baseline specification.[9] Similarly to De Paola et al. (2022), we allow for differential trends in credit achievement in the first and second semester. Including such effects is useful to account for some peculiarities of the Italian context. In the Italian university system, students have multiple exam sessions throughout the academic year (six sessions at UniMI). Typically, students who fail exams in the first semester can retake them in the second semester of the same academic year (but also in the following academic years). Passing students can also retake exams if they are not happy with their grades (they can "refuse" grades), which vary in the 18 (minimum passing grade) to 30 *with laude* range. Using a longitudinal setting in which all students are observed both before and after the pandemic allows us to tackle potential problems related to the imbalance of student characteristics across student cohorts, which is an issue potentially faced by studies implementing pooled cross-section difference-in-differences designs. Unlike De Paola et al. (2022), which only retains in the estimation sample students who did not drop out, we impute zero credits to the semesters following student drop-out in order to address potential issues of non-random drop-out with respect to student progression and COVID-19.[10]

The main identifying assumption of our model is that in the absence of COVID-19, the second vs. first semester gender gap contrast in student progression for the academic year 2019/20 would

[8] Rigorously speaking, model (1) is not a triple-difference model, since there is not a treated and an untreated group in the post period, rather, all students are (potentially) differentially treated according to their gender, with α_5 capturing the differential effect of the pandemic between genders. Thus, the model is more of a "before-after" model with heterogeneous effects than a DDD model. Yet, the model in equation (1) formally involves three differences, the first between the pre- and the post-COVID-19 academic years, the second between the first and the second semesters, and the third between male and females students; for ease of exposition, we will refer to it as a DDD model. Similar definitions are given by several related papers (e.g., De Paola et al., 2022; Rodríguez-Planas, 2022), which implement a research design similar to the one in this paper (omitting the triple difference by gender) and define it a "difference-in-differences"-like model.

[9] De Paola et al. (2022) include student fixed effects only in some specifications to check the sensitivity of results.

[10] Indeed, COVID-19 might have also affected student drop-out, inducing a sample selection bias. In other words, if dropouts are excluded from the analysis, only relatively better performing students, or students resilient to COVID-19, may remain in the estimation sample. Omitting from the sample students who might have partly dropped out from university because of the pandemic is likely to lead to an underestimation of the negative effect of COVID-19 on students' academic progression. The different time at risk of dropout is captured by the student fixed effects, which in turn absorb the student cohort fixed effects. We primarily focus on credits earned because, due to the possibility of sitting exams in several exam sessions in different semesters and academic years, semester GPAs are computed on a different number of exams for different students. Notwithstanding this caveat, later in the paper we also present results for semester GPAs. See Bonaccolto-Töpfer and Castagnetti (2021) for an extensive analysis of the effect of COVID-19 on the average grades of students at the University of Pavia (Northern Italy).

have remained the same as in the previous years. In order to test for this pre-COVID parallel trend assumption in the triple contrast, we also estimate the following event-study version of the DDD model (cf. Rodríguez-Planas, 2022):

$$\begin{aligned} y_{ijt} = \alpha_0 + \alpha_1 sem2_t + \sum_j \alpha_{2,j}(sem2_t \times \mathbb{1}(t \in S)) + \alpha_3(sem2_t \times female_i) \\ + \sum_j \alpha_{4,j}(female_i \times \mathbb{1}(t \in S)) + \sum_j \alpha_{5,j}(sem2_t \times female_i \times \mathbb{1}(t \in S)) + a.y._j + u_i + \epsilon_{ijt}, \end{aligned} \tag{2}$$

where $j = 2017, 2019$, while 2018 is the reference academic year (the year before the onset of COVID-19), and $\mathbb{1}(.)$ is the indicator function. Periods $t = 1, 3, 5$ correspond to first semesters, and periods $t = 2, 4, 6$ to second semesters. The set $S = \{2, 6\}$ is therefore the set of periods corresponding to the second semesters for which interaction terms with academic years are included (as period $t = 4$ belongs to the reference academic year, 2018). This specification allows for the coefficients in all double and triple interactions involving terms that depend on academic years to be year-specific. If the pre-pandemic parallel trend assumption holds, we should have $\alpha_{5,j} = 0$ for $j = 2017$, i.e., these triple interactions should not differ statistically in the pre-pandemic academic years (2017 and the baseline year 2018).

Finally, we carry out an heterogeneous analysis by broad college major, and estimate the following model:

$$
\begin{aligned}
y_{ijmt} =& \alpha_0 + \alpha_1\, sem2_t + \sum_{k \neq r} \alpha_{1,k} (sem2_t \times \mathbb{1}(k = m)) \\
& + \alpha_2 (sem2_t \times a.y.\ 2019/20_j) + \sum_{k \neq r} \alpha_{2,k} (sem2_t \times a.y.\ 2019/20_j \times \mathbb{1}(k = m)) \\
& + \alpha_3 (sem2_t \times female_i) + \sum_{k \neq r} \alpha_{3,k} (sem2_t \times female_i \times \mathbb{1}(k = m)) \\
& + \alpha_4\, a.y.\ 2019/20_j + \sum_{k \neq r} \alpha_{4,k} (a.y.\ 2019/20_j \times \mathbb{1}(k = m)) \\
& + \alpha_5 (female_i \times a.y.\ 2019/20_j) + \sum_{k \neq r} \alpha_{5,k} (female_i \times a.y.\ 2019/20_j \times \mathbb{1}(k = m)) \\
& + \alpha_6 (sem2_t \times female_i \times a.y.\ 2019/20_j) + \sum_{k \neq r} \alpha_{6,k} (sem2_t \times female_i \times a.y.\ 2019/20_j \times \mathbb{1}(k = m)) \\
& + u_i + \epsilon_{ijmt}, \qquad (3)
\end{aligned}
$$

where m is an integer indicating the college major in which the individual is enrolled and r the omitted (reference) college major category.

3 Institutional setting and data

Since 2000, the Bologna process (which introduced the so-called "3+2 reform") has organized the Italian university system around two main degree levels: (1) the first-level or bachelor's degree (*Laurea Triennale)*, whose legal duration is three years, and (2) the second-level or master's degree (2 further years) (*Laurea Magistrale).* For some specific fields such as law and medicine, there is a third type of degree structured as a single cycle (*Lauree Magistrali a Ciclo Unico*) lasting 5 or 6 years, depending on the major. We restrict the analysis to the student cohorts enrolled in the 2017/18, 2018/19, and 2019/20 academic years (three cohorts). Moreover, we focus on first-level and single-cycle degrees (i.e., first-time entrants in higher education), the legal durations of which are three and five (or six) years, respectively.[11] During our estimation period, different student

[11] Thus, all students are observed within the legal duration of their degrees and excluding the years exceeding the legal duration, the so-called Italian *fuori corso*.

Table 1: Structure of the data by academic year and year of enrollment

student cohort	academic year 2017/18	academic year 2018/19	academic year 2019/20
2017	1	2	3
2018		1	2
2019			1

Note. The numbers in each cell represent the year of study for each student cohort in each academic year.

cohorts are observed in different years of study in each academic year. For both first-level and single-cycle degrees, during the estimation period we can observe students in year 1 to year 3 of their degrees. The third year corresponds to the final year (according to the legal duration) of first-level degrees, while in single-cycle degrees students will have two or three more years of study (according to the legal duration of their degree). Table 1 reports the structure of the data. Students from the 2017/18, 2018/19, and 2019/20 cohorts are observed for years 1, 2, and 3; years 1 and 2; and year 1 of their studies, respectively. As mentioned in the previous section, we impute zero credits to the semesters following student drop out. All students are observed both before and after the first wave of COVID-19. Table 1 shows how the effect of COVID-19 on student progression is identified using students in their third, second, and first year of studies, which, of course, belong to different student cohorts.

In Italy, students sit exams in several sessions. At the university of Milan, students are granted six exam sessions per year. Exam sessions are scheduled throughout the year, depending on the semester in which the courses are taught. The first session shortly follows the end of each course. Moreover, students can refuse grades if they pass an exam but are not satisfied with their performance.[12] For these reasons (exam resits due to fails or refused grades), students usually earn more credits in the second semester of each academic year compared to the first semester. Exams generally award 3, 6, 9, or 12 credits. One credit corresponds to 25 hours of student workload. One academic year generally corresponds to 60 credits, split into two semesters. The workload for each

[12] This phenomenon may partly contribute to many students graduating beyond the legal (or regular) degree duration in Italy (Garibaldi et al., 2012).

semester tends to be balanced, but it is not exactly the same. Three-year degrees correspond to 180 credits and single-cycle degrees to 360 credits for medicine and dentistry and 300 credits for other majors.

Figure 1 plots the average credits earned and confidence intervals for each semester in the period spanning the 2017/18-2019/20 academic years by gender.[13] The average for the second semester is always above that of the first semester, confirming that semesters are not all alike. For this reason, our regression models (like De Paola et al., 2022) control for indicators for the first and second semester. Student progression seems to be more gender-balanced in the first semesters than in the second semesters, in which there is an advantage for women. In the raw data, at least visually, the gender gap appears to be rather stable in the second semesters and increases over time for the first semesters.

Figure 2 shows the distribution of student credits by semester for the whole period, considering all semesters, and for the first and second semesters separately. The data show some bunching, owing to the fact that credits are earned in bunches, corresponding to exams passed and registered, and a mass point at zero credits accounting for about 30% of observations.[14]

4 Results

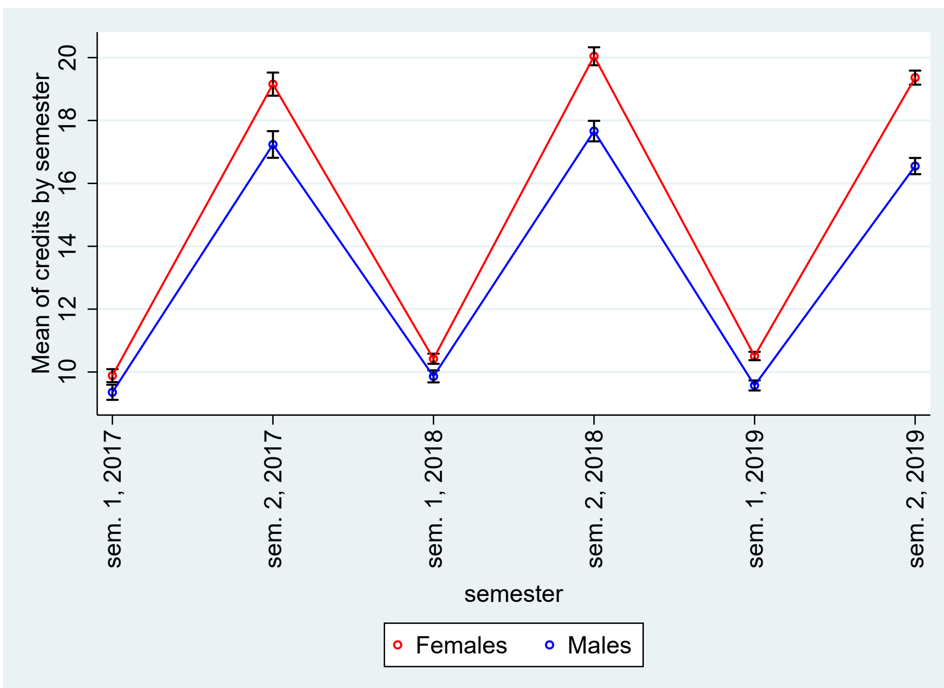
4.1 Credits earned

Our baseline DDD results for the number of earned credits are reported in column (1) of Table 2. The DDD estimates show that, on average, before the pandemic students earned about 8 more credits in the second semester than in the first. A similar result is found by De Paola et al. (2022) for the University of Calabria. This finding can be explained both by some "learning by doing", especially for freshmen, as university studies require a good amount of self-organization compared to secondary education, or as previously mentioned, by some features of the Italian higher education system (e.g., the possibility of re-sitting exams several times during the year and of refusing grades).

[13] Means and confidence intervals are computed on our estimation sample.

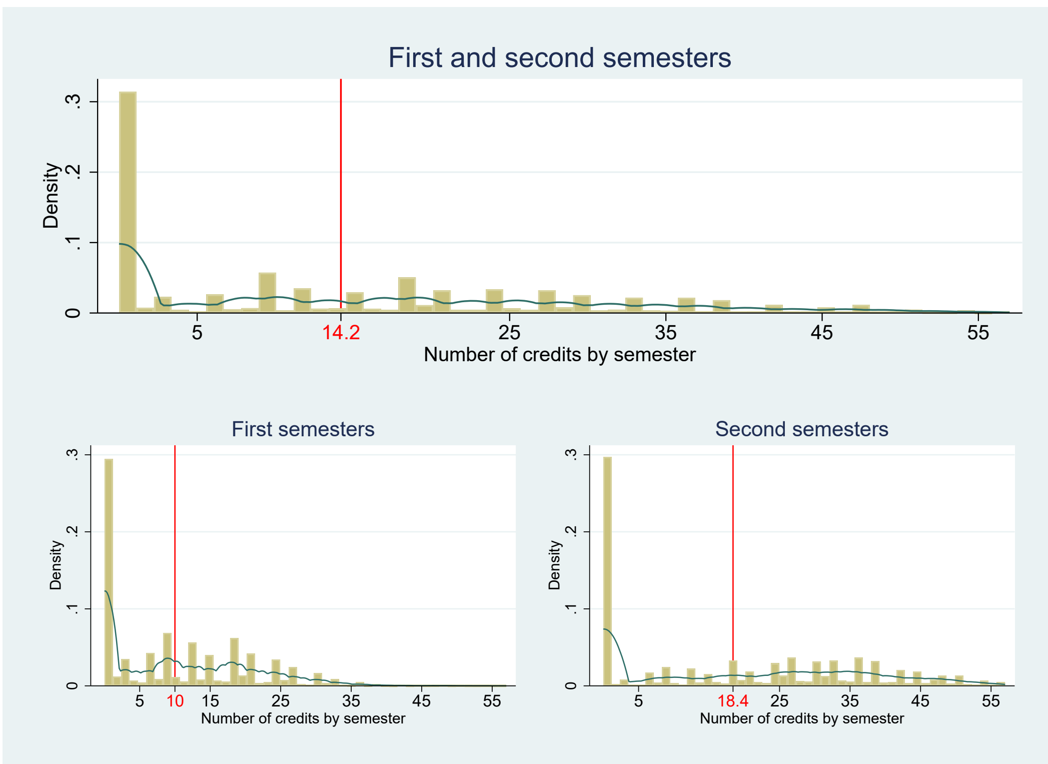
[14] The average GPA by gender and the distribution of GPA by semester and gender are reported in Figure A2 and Figure A3 in the online Appendix, respectively.

Figure 1: Means and confidence intervals of credits by semester and student gender



Note. Mean credits earned by students in each semester of the 2017/18-2019/20 academic years, by gender. Bars represent 95% confidence intervals. Means and confidence intervals are computed on the estimation sample.

Figure 2: Distribution of credits by semester



Note. Distribution of credits by semester for the 2017/18-2019/20 academic years — all semesters and first and second semesters separately. The vertical lines represent the means. Distributions and means are computed on the estimation sample.

Table 2: DDD estimates of the number of earned credits and GPA by semester

VARIABLES	(1) credits	(2) credits (ES)	(3) GPA	(4) GPA (ES)
sem2	7.829***	7.802***	-0.005	0.030
	(0.875)	(0.889)	(0.107)	(0.096)
sem2 × a.y. 2017/18		0.080		-0.107
		(0.412)		(0.154)
sem2 × a.y. 2019/20	-0.851**	-0.824**	0.222*	0.187*
	(0.347)	(0.357)	(0.113)	(0.100)
sem2 × female	1.675***	1.819***	0.093	0.090
	(0.601)	(0.623)	(0.088)	(0.088)
female × a.y. 2017/18		-0.008		-0.149
		(0.318)		(0.116)
female × a.y. 2019/20	0.635***	0.642***	0.227**	0.186**
	(0.207)	(0.182)	(0.086)	(0.083)
sem2 × female × a.y. 2017/18		-0.430		0.012
		(0.541)		(0.155)
sem2 × female × a.y. 2019/20	0.201	0.058	0.056	0.059
	(0.341)	(0.400)	(0.069)	(0.071)
a.y. 2017/18	-0.313	-0.225	-0.181**	-0.039
	(0.240)	(0.319)	(0.069)	(0.113)
a.y. 2019/20	-0.284	-0.260	0.124	0.166*
	(0.201)	(0.196)	(0.093)	(0.090)
Constant	10.073***	10.045***	25.641***	25.624***
	(0.385)	(0.382)	(0.064)	(0.059)
Observations	157,252	157,252	90,541	90,541
R-squared	0.688	0.688	0.622	0.622
Effect for females	-0.650	-0.767	0.278**	0.245**
	(0.419)	(0.491)	(0.135)	(0.117)
Mean baseline (female, sem2)	19.75	20.04	25.67	25.73
% effect females	-3.292	-3.825	1.082	0.952
Effect for males	-0.851**	-0.824**	0.222*	0.187*
	(0.347)	(0.357)	(0.113)	(0.0997)
Mean baseline (male, sem2)	17.52	17.66	25.22	25.24
% effect males	-4.860	-4.668	0.880	0.739
Differential effect for females	0.201	0.0577	0.0558	0.0585
	(0.341)	(0.400)	(0.0692)	(0.0710)

Note. Standard errors clustered at the degree-course level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is the number of credits earned by semester in columns (1) and (2) and GPA in columns (3) and (4). DDD (event study-DDD) models are reported in columns (1) and (3) ((2) and (4)). All models include individual fixed effects. The estimation sample includes students belonging to the 2017/18, 2018/19, and 2019/20 cohorts. The effects for males and females are α_2 and $\alpha_2 + \alpha_5$ from model (1), respectively, in columns (1) and (3). The effects for males and females are $\alpha_{2,2019}$ and $\alpha_{2,2019} + \alpha_{5,2019}$, respetively, from model (2) in columns (2) and (4).

The DDD estimates point to a negative effect of COVID on the amount of earned credits, with a point estimate of -0.9 credits for men (about -5% at the baseline), statistically significant at the 5% level, while the point estimate for females is lower, at -0.7, and is not statistically significant. Yet, given the similarity of the point estimates for men and women, the null hypothesis of equal effects by gender cannot be rejected at conventional significance levels (see the bottom of Table 2). Column (2) reports the results of the event study-DDD estimates, which support the validity of the parallel trend assumption in that the triple interaction ($sem2 \times female \times a.y.$ 2017/18) is not statistically significant at conventional levels.[15] Our estimates are lower than those reported by De Paola et al. (2022), who estimate a negative effect of online teaching of about -1.4 credits for the University of Calabria.

Table 1, which presents the structure of the data, shows a peculiarity of the 2019/20 student cohort, for which (unlike previous cohorts) the contrast between the number of credits in the first and second semesters can be computed only for one academic year, the one affected by COVID-19. For this reason, we checked the robustness of the estimates to omitting the cohort enrolled in the 2019/20 academic year (i.e. freshmen). The point estimates of the triple-difference coefficient (the complete set of estimates is available upon request) are quite small, at 0.2 (p-value = 0.66) for credits earned and 0.002 for GPA (p-value = 0.98), and are very similar to those reported in columns (1) and (3) of Table 2.

4.2 Credits earned by broad college major

Women are under-represented in STEM (Bottia et al., 2015; Kahn and Ginther, 2017; Carlana and Fort, 2022) and are less likely to persist in STEM degrees (Griffith, 2010). Thus, it is important to investigate whether the pandemic contributed to worsening the gender gap in STEM majors in particular. Indeed, negative shocks on student progression or grades (Rask, 2010) and worsening performance compared to men (Kugler et al., 2017) may reduce female students' persistence in STEM majors and discourage future student cohorts from choosing to pursue STEM degrees.

In this subsection, we present an effect heterogeneity analysis by college major. In particular,

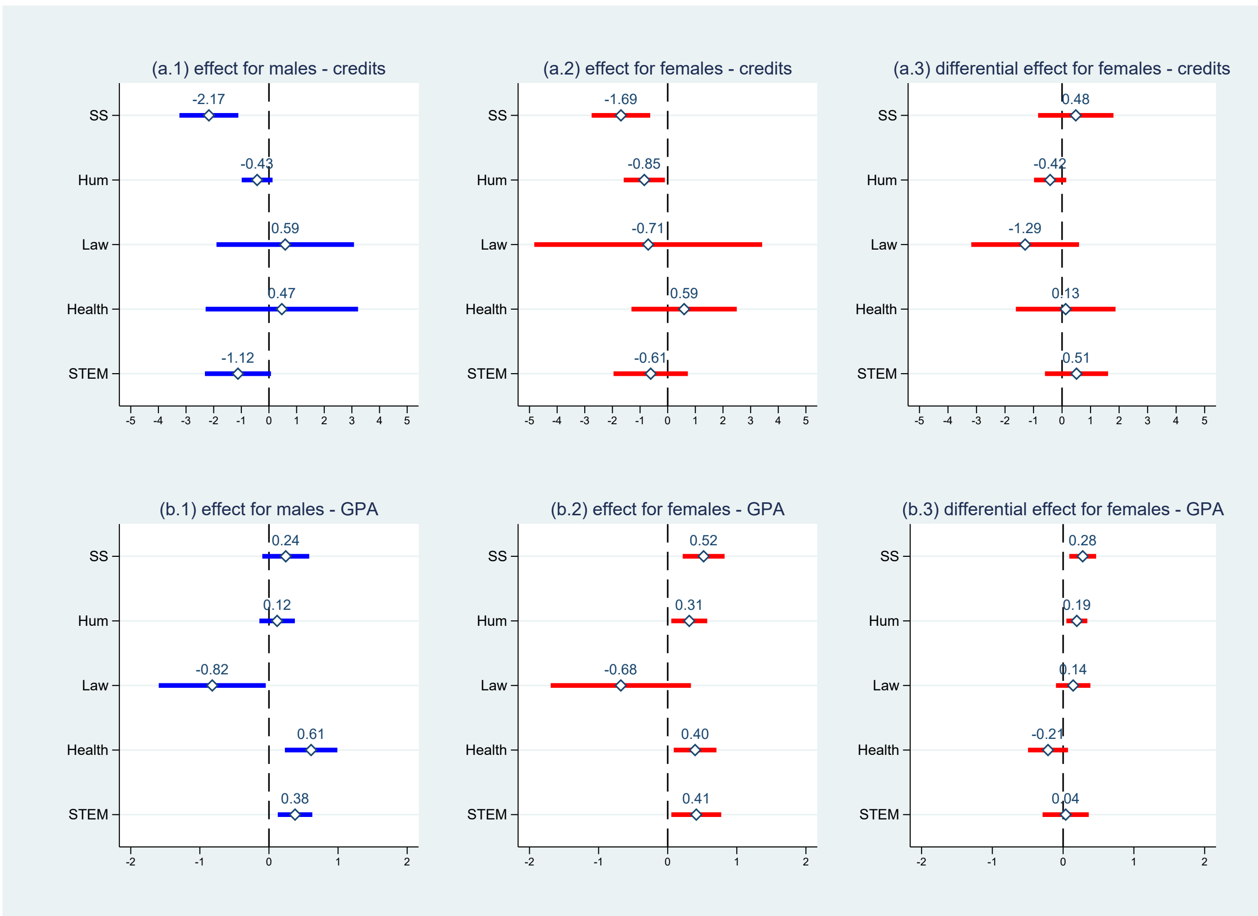
[15] Point estimates and confidence intervals for the triple interactions are plotted in Figure A1 in the online Appendix.

we estimate the model in equation (3), which also includes interaction terms with college majors. In order to retain enough statistical power, degrees are grouped into five broad majors: social sciences, humanities, law, health, and STEM. This analysis is informative as online learning may be differentially effective depending on the nature of the material being taught (Contini et al., 2021). More quantitative courses are generally taught in STEM compared to social sciences and health, with law and the humanities being the least quantitative majors. Graphs (a.1) and (a.2) of Figure 3 show the effect of COVID-19 on men and women, respectively, while graph (a.3) displays the difference in the coefficient estimates between women and men. The coefficient estimates are reported in Table A1 in the online Appendix. First, COVID-19 did not negatively impact student progression in all majors. Significant negative effects are estimated in social sciences, for both women (-1.7 credits, i.e., -4.9%) and men (-2.2 credits, i.e., -2.5%), in the humanities for women (-0.8 credits, i.e., -4.9%), and in STEM for men (a fall of about one credit, i.e., -1.9%). However, gender differences are generally low (except in law, for which they are quite imprecisely estimated, however) and never statistically significant. All in all, the pandemic does not seem to have particularly impacted quantitative majors or women within quantitative majors.

4.3 GPA

The number of earned credits is not the only student performance indicator (PI) that could be considered, although it is one on which the Italian Higher Education Quality Assurance (QA) system places significant weight. For instance, the percentages of first-year students earning at least 20 or 40 credits per year represent two important PIs defined by the National Agency of Evaluation of the University System and Research (ANVUR) on which bachelor and master degrees are regularly assessed (Bratti et al., 2022). Another important measure of student performance is GPA. For this reason, in this section we comment on the results of the DDD analysis using GPA as the dependent variable. Interestingly, column (3) of Table 2 shows that the reduction of credits was partly compensated by a higher GPA, by about 0.2 points for men (an increase of about 0.9%) and 0.3 points for women (1%). This result mimics qualitatively the finding of Rodríguez-Planas (2022) of a positive effect of COVID-19 on GPA. As expected, the effect on GPA in the Italian

Figure 3: DDD estimates of credits earned and GPA by broad college major



Note. The figure shows plots of the effect of the pandemic (with 95% confidence intervals) by gender and broad college major (estimated using interactions) in terms of number of credits, in graphs (a.1)-(a.2), and GPA in graphs (b.1)-(b.2). Graphs (a.3) and (b.3) display the female-male contrasts for credits earned and GPA, respectively. Point estimates obtained from model (3) and standard errors are also reported in the bottom part of Table A1 in the online Appendix.

context is lower than the 5% effect size reported by Rodríguez-Planas (2022). The reason is that in Italy, an institutional mechanism that allows students to retain some control over their GPA (the possibility of refusing grades) was in place both before and after COVID-19. In contrast, flexible grading was introduced by the institution studied in Rodríguez-Planas (2022) only in the post-COVID-19 period. For Italy, a possible explanation for the positive effect of COVID-19 on GPA is that during the first wave of the pandemic not only teaching but also exams were organized remotely. Online exams might have reduced the cost of sitting or attempting exams (e.g., lower travel costs for students residing far from the university) and, accordingly, also the cost of refusing grades. This could explain the simultaneous decrease in credits and increase in semester GPA.[16]

4.4 GPA by broad college major

Analogously to credits earned, we estimate the gendered impact of COVID-19 by broad college major on GPA. Results are shown in the bottom panel of Figure 3, while point estimates are reported in Table A2 in the online Appendix. Overall, COVID-19 had a positive impact on both males' and females' GPA in all fields but law. However, the coefficients are not always statistically significant. More specifically, women show a significant and positive effect in the social sciences (0.5, i.e., 2.1%) and the humanities (0.3, i.e., 1.2%), while in health and STEM a positive effect is found for both women (0.4, i.e., 1.5%, and 0.4, i.e., 1.7%, respectively) and men (0.6, i.e., 2.3%, and 0.4, i.e., 1.5%, respectively). As for law, we find a negative but non-significant effect for women (−0.7, i.e., −2.6%) and a negative and statistically significant one for men (−0.8, i.e., −3.1%). The effects in law represent the strongest effects among all majors, both in terms of percentage and in absolute terms, but are very imprecisely estimated. Finally, gender differences are statistically significant (at the 5% level) for the social sciences and the humanities. In both fields, women have improved their relative performance in terms of GPA (by 0.3 and 0.2 points, respectively).

[16] Unfortunately, we are not able to estimate the effect on the number of exam attempts, failed exams, or refused results, as some teachers do not record exam fails (counting students as absent), while others do, and refused grades do not enter student records.

5 Concluding remarks

There are concerns that, like for other domains such as employment and income, women might have also shouldered the brunt of the COVID-19 crisis in education, for instance, because of the higher impact of COVID-19 on women's mental health or family workloads. Using data on the entire population of students at a large public Italian university located in the Italian region most affected by the first pandemic and a triple-difference identification strategy, we find that COVID-19 negatively impacted students' academic progression (credits earned) but was not more harmful for female students. We also observe a positive effect of COVID-19 on semester GPA for both men and women. The analysis of gender differences by broad college major shows that the pandemic did not particularly harm performance in quantitative degrees (STEM) or women enrolled in quantitative degrees. In contrast, women's GPA improved compared to men in social sciences and humanities degrees.

References

Agostinelli, F., M. Doepke, G. Sorrenti, and F. Zilibotti (2022). When the great equalizer shuts down: Schools, peers, and parents in pandemic times. *Journal of Public Economics 206*, 104574.

Alon, T., M. Doepke, J. Olmstead-Rumsey, and M. Tertilt (2020). This time it's different: The role of women's employment in a pandemic recession. Technical report, National Bureau of Economic Research.

Aucejo, E. M., J. French, M. P. Ugalde Araya, and B. Zafar (2020, November). The impact of COVID-19 on student experiences and expectations: Evidence from a survey. *Journal of Public Economics 191*, 104271.

Bonaccolto-Töpfer, M. and C. Castagnetti (2021). The COVID-19 pandemic: A threat to higher education? Technical report, Diskussionspapiere.

Bottia, M. C., E. Stearns, R. A. Mickelson, S. Moller, and L. Valentino (2015). Growing the roots of STEM majors: Female math and science high school faculty and the participation of students in STEM. *Economics of Education Review 45*, 14–27.

Bratti, M., G. Barbato, D. Biancardi, C. Conti, and M. Turri (2022). Degree-level determinants of university student performance. In D. Checchi and J. T. (Eds.), *Teaching, Research and Academic Careers*, Chapter 10. Springer.

Carlana, M. and M. Fort (2022). Hacking gender stereotypes: Girls' participation in coding clubs. *AEA Papers and Proceedings 112*, 583–87.

Casalone, G., A. Michelangeli, J. Östh, and U. Türk (2021). The effect of lockdown on students' performance: A comparative study between Sweden, Italy, and Turkey. *University of Milan Bicocca Department of Economics, Management and Statistics Working Paper* (473).

Contini, D., M. L. Di Tommaso, C. Muratori, D. Piazzalunga, and L. Schiavon (2021). The COVID-19 pandemic and school closure: Learning loss in mathematics in primary education. IZA Discussion Papers 14785, Institute of Labor Economics (IZA).

De Paola, M., F. Gioia, and V. Scoppa (2022). Online teaching, procrastination and students' achievement: Evidence from COVID-19 induced remote learning. IZA Discussion Papers 15031, Institute of Labor Economics (IZA).

Del Boca, D., N. Oggero, P. Profeta, and M. Rossi (2020). Women's and men's work, housework and childcare, before and during COVID-19. *Review of Economics of the Household 18*(4), 1001–1017.

Delaney, J. M. and P. J. Devereux (2021). High school rank in math and English and the gender gap in STEM. *Labour Economics 69*, 101969.

Engzell, P., A. Frey, and M. D. Verhagen (2021). Learning loss due to school closures during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences 118*(17).

Etheridge, B. and L. Spantig (2020). The gender gap in mental well-being during the Covid-19 outbreak: Evidence from the UK. Technical report, ISER Working paper series.

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

Giurge, L. M., A. V. Whillans, and A. Yemiscigil (2021). A multicountry perspective on gender differences in time use during COVID-19. *Proceedings of the National Academy of Sciences 118*(12).

Graeber, D., A. S. Kritikos, and J. Seebauer (2021). COVID-19: A crisis of the female self-employed. *Journal of Population Economics 34*(4), 1141–1187.

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

Gruber, J. (1994). The incidence of mandated maternity benefits. *The American Economic Review 84*(3), 622–641.

Kahn, S. and D. Ginther (2017, June). Women and STEM. Working Paper 23525, National Bureau of Economic Research.

Kofoed, M. S., L. Gebhart, D. Gilmore, and R. Moschitto (2021). Zooming to class?: Experimental evidence on college students' online learning during COVID-19. IZA Discussion Papers 14356, Institute of Labor Economics (IZA).

Kugler, A. D., C. H. Tinsley, and O. Ukhaneva (2017, August). Choice of majors: Are women really different from men? Working Paper 23735, National Bureau of Economic Research.

Kuhfeld, M., J. Soland, B. Tarasawa, A. Johnson, E. Ruzek, and J. Liu (2020). Projecting the potential impact of COVID-19 school closures on academic achievement. *Educational Researcher 49*(8), 549–565.

Liu, X., W. He, L. Zhao, and J.-C. Hong (2021). Gender differences in self-regulated online learning during the COVID-19 lockdown. *Frontiers in Psychology 12*.

Meelissen, M. R. and M. Drent (2008). Gender differences in computer attitudes: Does the school matter? *Computers in Human Behavior 24*(3), 969–985.

Olden, A. and J. Møen (2020). The triple difference estimator. Technical Report 2020/1, NHH Dept. of Business and Management Science Discussion Paper.

Orlov, G., D. McKee, J. Berry, A. Boyle, T. DiCiccio, T. Ransom, A. Rees-Jones, and J. Stoye (2021). Learning during the COVID-19 pandemic: It is not who you teach, but how you teach. *Economics Letters 202*, 109812.

Prowse, R., F. Sherratt, A. Abizaid, R. L. Gabrys, K. G. Hellemans, Z. R. Patterson, and R. J. McQuaid (2021). Coping with the COVID-19 pandemic: Examining gender differences in stress and mental health among university students. *Frontiers in Psychiatry 12*, 439.

Rask, K. (2010). Attrition in STEM fields at a liberal arts college: The importance of grades and pre-collegiate preferences. *Economics of Education Review 29*(6), 892–900.

Rodríguez-Planas, N. (2022). COVID-19, college academic performance, and the flexible grading policy: A longitudinal analysis. *Journal of Public Economics*, 104606.

Sevilla, A. and S. Smith (2020). Baby steps: The gender division of childcare during the COVID-19 pandemic. *Oxford Review of Economic Policy 36*(Supplement_1), S169–S186.

Supriya, K., C. Mead, A. D. Anbar, J. L. Caulkins, J. P. Collins, K. M. Cooper, P. C. LePore, T. Lewis, A. Pate, R. A. Scott, et al. (2021). Undergraduate biology students received higher grades during COVID-19 but perceived negative effects on learning. *Frontiers in Education 6*(759624).

Zamarro, G. and M. J. Prados (2021). Gender differences in couples' division of childcare, work and mental health during COVID-19. *Review of Economics of the Household 19*(1), 11–40.

A Online Appendix

Table A1: Earned credits by broad college major

Variables	(1) Social Sciences (ref.)	(2) Humanities	(3) Law	(4) Health	(5) STEM
sem2	12.64***	-8.980***	-3.032	-2.928*	-4.881***
	(0.740)	(0.997)	(3.800)	(1.488)	(1.464)
sem2 × a.y. 2019/20	-2.173***	1.745***	2.762**	2.640*	1.052
	(0.535)	(0.604)	(1.359)	(1.484)	(0.804)
sem2 × female	1.266***	1.662**	1.997**	1.436*	-1.672**
	(0.342)	(0.760)	(0.875)	(0.771)	(0.801)
sem2 × female × a.y. 2019/20	0.481	-0.899	-1.775	-0.353	0.0257
	(0.664)	(0.722)	(1.157)	(1.100)	(0.866)
Constant	9.989***	9.989***	9.989***	9.989***	9.989***
	(0.304)	(0.304)	(0.304)	(0.304)	(0.304)
Observations	157,252	157,252	157,252	157,252	157,252
R-squared	0.699	0.699	0.699	0.699	0.699
Effect for females	-1.692***	-0.846**	-0.705	0.594	-0.615
	(0.532)	(0.374)	(2.070)	(0.956)	(0.675)
Mean baseline (female, sem2)	26.12	17.17	22.12	22.12	16.62
% effect females	-4.927	-4.927	-4.927	-3.797	-3.797
Effect for males	-2.173***	-0.428	0.589	0.466	-1.122*
	(0.535)	(0.280)	(1.249)	(1.384)	(0.600)
Mean baseline (male, sem2)	23.80	13.73	17.44	17.44	16.82
% effect males	-2.491	-2.491	-2.491	-1.920	-1.920
Differential effect for females	0.481	-0.418	-1.294	0.128	0.507
	(0.664)	(0.284)	(0.948)	(0.877)	(0.556)

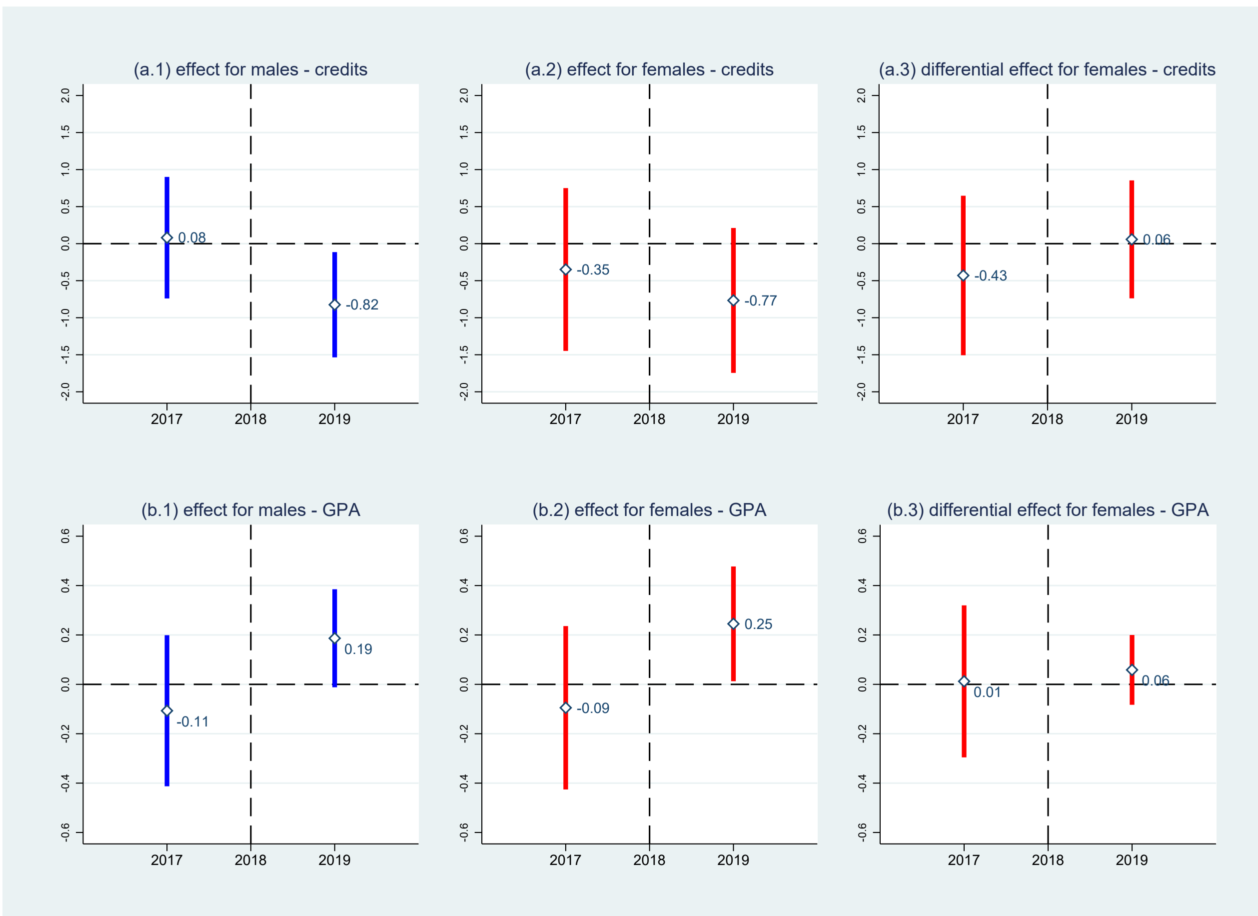
Note. Standard errors clustered by degree course in parentheses. *** p<0.01, ** p<0.05, * p<0.1. *sem2* is an indicator for the second semester, *female* an indicator for female students, and *a.y. 2019/20* is an indicator for the 2019/20 academic year. All models include individual fixed effects. The estimation sample includes students belonging to the 2017/18, 2018/19, and 2019/20 cohorts. For men, the effect for the reference college major is the coefficient α_2, while the coefficients for the other majors are $\alpha_2 + \alpha_{2,k}$ estimated from model (3). For women, the effect for the reference college major is the coefficient $\alpha_2 + \alpha_6$, while the coefficients for the other majors are $\alpha_2 + \alpha_6 + \alpha_{2,k} + \alpha_{6,k}$ estimated from model (3).

Table A2: GPA by broad college major

Variables	(1) Social Sciences (ref.)	(2) Humanities	(3) Law	(4) Health	(5) STEM
sem2	-0.254	0.0537	0.967**	0.112	0.382
	(0.257)	(0.305)	(0.473)	(0.320)	(0.281)
sem2 × a.y. 2019/20	0.243	-0.125	-1.063**	0.367	0.135
	(0.171)	(0.214)	(0.424)	(0.256)	(0.212)
sem2 × female	0.131	0.0400	-0.131	0.118	-0.145
	(0.210)	(0.268)	(0.220)	(0.233)	(0.247)
sem2 × female × a.y. 2019/20	0.277***	-0.0833	-0.135	-0.491***	-0.241
	(0.0952)	(0.121)	(0.155)	(0.171)	(0.190)
Constant	25.59***	25.59***	25.59***	25.59***	25.59***
	(0.0659)	(0.0659)	(0.0659)	(0.0659)	(0.0659)
Observations	90,541	90,541	90,541	90,541	90,541
R-squared	0.624	0.624	0.624	0.624	0.624
Effect for females	0.520***	0.312**	-0.678	0.397**	0.414**
	(0.152)	(0.130)	(0.509)	(0.155)	(0.181)
Mean baseline (female, sem2)	25.27	26.27	26.31	26.31	24.86
% effect females	2.058	1.188	-2.578	1.507	1.665
Effect for males	0.243	0.118	-0.820**	0.610***	0.378***
	(0.171)	(0.129)	(0.389)	(0.191)	(0.125)
Mean baseline (male, sem2)	24.53	26.40	25.99	25.99	24.75
% effect males	0.961	0.450	-3.118	2.319	1.520
Differential effect for females	0.277***	0.194**	0.142	-0.213	0.0360
	(0.0952)	(0.0744)	(0.122)	(0.142)	(0.164)

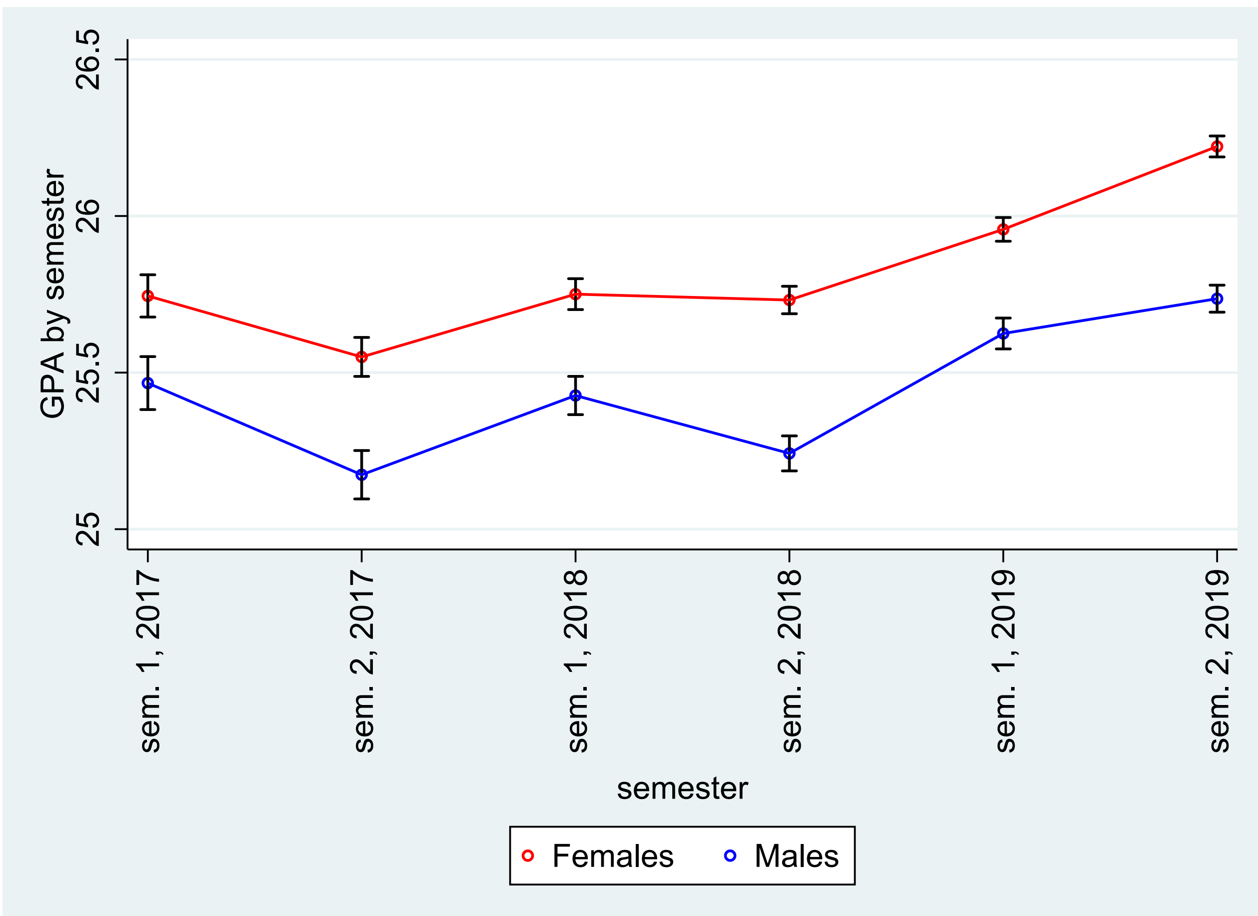
Note. Standard errors clustered by degree course in parentheses. *** p<0.01, ** p<0.05, * p<0.1. *sem2* is an indicator for the second semester, *female* an indicator for female students, and *a.y. 2019/20* is an indicator for the 2019/20 academic year. All models include individual fixed effects. The estimation sample includes students belonging to the 2017/18, 2018/19, and 2019/20 cohorts. For men, the effect for the reference college major is the coefficient α_2, while the coefficients for the other majors are $\alpha_2 + \alpha_{2,k}$ estimated from model (3). For women, the effect for the reference college major is the coefficient $\alpha_2 + \alpha_6$, while the coefficients for the other majors are $\alpha_2 + \alpha_6 + \alpha_{2,k} + \alpha_{6,k}$ estimated from model (3).

Figure A1: Event-study estimates of credits earned and GPA



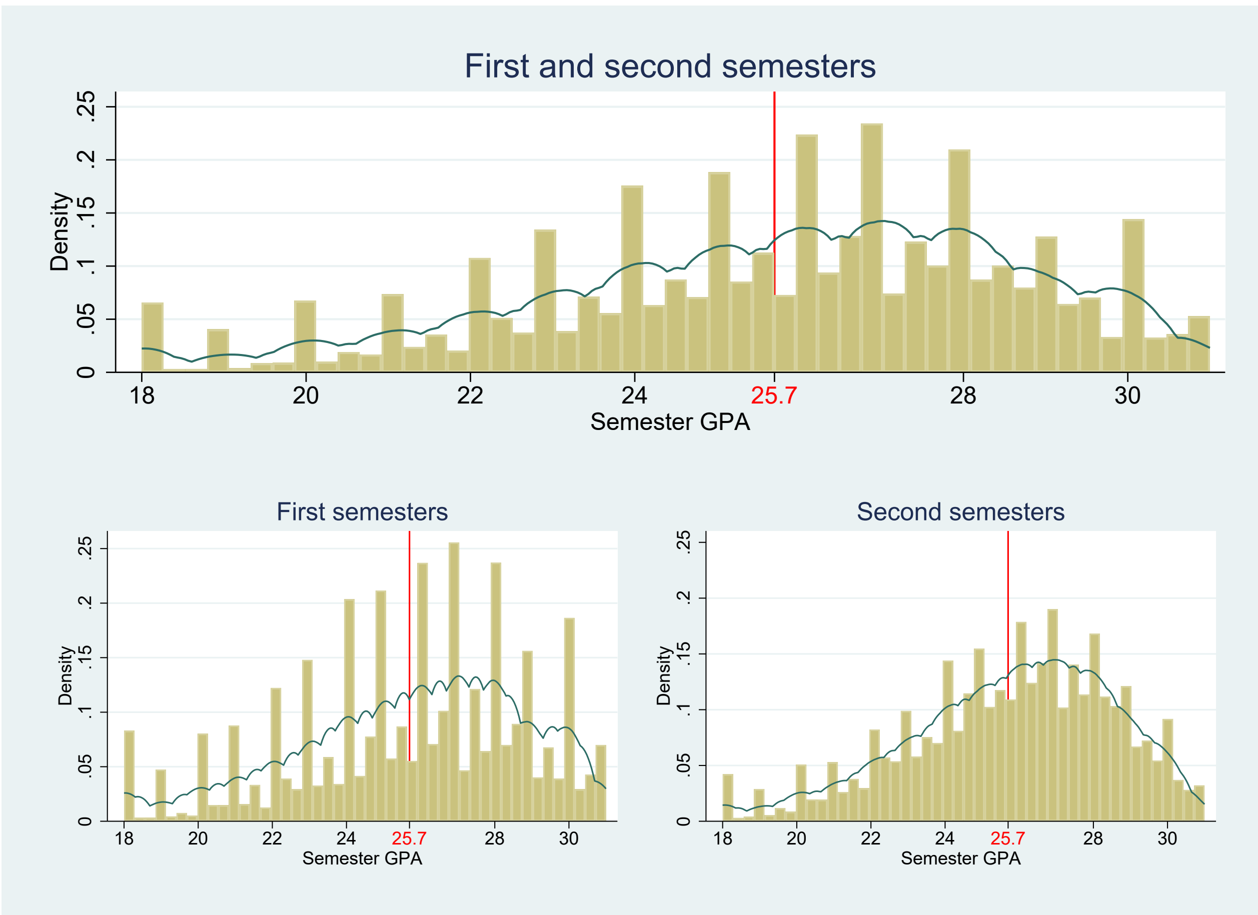
Note. The figures show plots of the coefficients (with 95% confidence intervals) for the second semester by female interactions ($\alpha_{5,j}$ in model (2)) for the number of credits earned in graphs (a.1)-(a.3) and GPA in graphs (b.1)-(b.3).

Figure A2: Means and confidence intervals of GPA by semester and student gender



Note. GPA in each semester of the 2017/18-2019/20 academic years by gender. Bars represent 95% confidence intervals. Means and confidence intervals are computed on the estimation sample.

Figure A3: Distribution of GPA by semester



Note. Distribution of GPA by semester in the 2017/18-2019/20 academic years, showing all semesters together and the first and second semesters separately. The vertical lines represent the means. Distributions and means are computed on the estimation sample. The maximum grade "30 *with laude*" is conventionally coded as 31.