# The COVID-19 pandemic: A threat to higher education?

Marina Bonaccolto-Töpfer
University of Erlangen-Nürnberg

Carolina Castagnetti
University of Pavia

(April 2021)

LASER Discussion Papers - Paper No. 125

(edited by A. Abele-Brehm, R.T. Riphahn, K. Moser and C. Schnabel)

Correspondence to:

Marina Bonaccolto-Töpfer, Lange Gasse 20, 90403 Nuremberg, Germany, Email: marina.toepfer@fau.de.

**Abstract**

Transition to online teaching during the first wave of the COVID-19 pandemic has led to various concerns about educational quality. So far, researchers have mainly focused on the effects on school teaching. This paper looks at the effects on a large Italian university (University of Pavia, Lombardy). Administrative data allows us to track both students' evaluation of teaching and student performance. Using a difference-in-differences design, we exploit the fact that the summer term 2020 started right after the first lockdown and compare students' outcome during this term to those of the same term in the previous year. In contrast to the literature, our results suggest no substantial effects of the pandemic on higher education. The findings are robust across various dimensions of courses, students and lecturers. In particular, the results suggest also no difference between top and bottom students or students from wealthier and poorer families.



**Zusammenfassung**

Der Übergang zur Online-Lehre während der ersten Welle der COVID-19-Pandemie hat zu Sorgen über die Qualität der Lehre geführt. Bisher hat sich die Forschung größtenteils auf Effekte der Pandemie auf Schulen konzentriert. Dieses Papier analysiert die Auswirkungen auf eine große italienische Universität (Universität Pavia, Lombardei). Administrative Daten erlauben es uns, sowohl die Evaluation der Lehre als auch die Leistung von Studierenden zu betrachten. Mittels eines Differenz-von-Differenzen Ansatzes verwenden wir die Tatsache, dass das Sommersemester 2020 direkt nach dem ersten Lockdown gestartet ist, und vergleichen Ergebnisse von Studierenden in diesem Semester mit denen im gleichen Semester des vorangegangenen Jahres. Anders als die Literatur finden wir keine substanziellen Effekte der Pandemie auf Hochschulbildung. Unsere Ergebnisse sind entlang diverser Kurs-, Studierenden- und Dozierenden-Dimensionen robust. Insbesondere suggerieren die Ergebnisse auch keine Unterschiede zwischen sehr guten und schwachen Studierenden oder zwischen Studierenden aus reicheren und ärmeren Familien.



**Author note**

We thank Paolo Bertoletti, Giulio Pedrini and Claus Schnabel for helpful comments and suggestions and in particular the University of Pavia for data provision.

# 1 Introduction

The first wave of the COVID-19 pandemic hit Italy in February 2020. One of the measures taken by the Italian government to counteract the spread of the COVID-19 pandemic was the closure of educational institutions from kindergartens to universities. On March 8, 2020 a decree issued by the President of the Council of Ministers suspended classes in all Italian educational institutions with the possibility to carry out distance learning activities such as online teaching. The North of Italy was hit particularly hard during the first COVID-19 wave in spring 2020. Therefore, several regions in the North such as Lombardy anticipated the closure of schools, kindergartens and universities to the end of February and thus also the transition to online teaching. Educational institutions have adopted online teaching and responded to the closure very differently. Primary and secondary schools generally reacted with a delay to the measures (closures and transition from face-to-face to online teaching) and reduced teaching hours,[1] while universities responded almost without any delay. In case of universities, the hours of teaching provided did not change, nor did the evaluation of exams. Consequently, the measures taken may have had different effects on schools and universities.

This paper analyzes various effects of the transition to online teaching due to the first wave of the COVID-19 pandemic on higher education at the University of Pavia, a large university in Lombardy (Italy).[2] We use administrative data from the University of Pavia for academic years 2018/2019 and 2019/2020. As the February-2020 closure coincided with the beginning of the second semester of the academic year 2019/2020, we can identify the causal impact of online teaching along two dimensions. First, we look at the effect on Student Evaluations of Teaching (SET), a commonly used measure for teaching quality (see Bertoni et al., 2020, and the references therein). Second, we investigate the effects of the pandemic on student performance in terms of grades, exam failure rates and graduation grades over time.

Most of the papers that have addressed the effects of the pandemic on higher education base their considerations on general and descriptive aspects without showing empirical evidence from university data (Bahasoan et al., 2020; Mishra et al.,

[1]The Ministry of Education has set minimum requirements in terms of hours to be offered. For example, for secondary schools this requirement has been set at 15 weekly hours compared to the previous 30 weekly hours offered.

[2]The University of Pavia, established in 1361, is one of the oldest universities in the world. It was the only university in Lombardy region until the end of the 19th century. The university has more than 20,000 students from Italy and all over the world.

2020; Susskind and Vines, 2020). An exception is provided by Aucejo et al. (2020) who use survey data and find pronounced negative effects of the COVID-19 pandemic on students' outcomes and expectations about the future that appear robust across various dimensions (such as family background). The results of Aucejo et al. (2020) suggest that both financial (such as lack of financial resources to complete studies) and health effects (e.g. fear of becoming sick) of the pandemic need to be addressed in order to circumvent rising inequality in higher education. Their findings are in line with other studies on the effects for students of recessions on future wages (Kahn, 2010) or graduation (Oreopoulos et al., 2012). Further, the results are also in line with those from the emerging COVID-19 literature looking at the effects of the pandemic on schools (Agostinelli et al., 2020; Chetty et al., 2020; Engzell et al., 2020; Kuhfeld et al., 2020). For example, Agostinelli et al. (2020) found that school closures had a large and presumably persistent effect on educational outcomes of high school students that is highly unequally distributed. High school students from poor neighborhoods suffered from a learning loss, while those from rich neighborhoods remained unaffected. Socioeconomic conditions such as family background appear to have contributed to growing educational inequality during the pandemic. Similarly, Chetty et al. (2020) found that student progress of an online program in Maths decreased in areas with poorer ZIP codes.

As mentioned above, so far, most studies on the emerging COVID-19 literature concentrated on schools (e.g. Bacher-Hicks et al., 2021; Engzell et al., 2020). To the best of our knowledge, this is the first paper that examines the effects of the COVID-19 pandemic on SET and student performance using the universe of all students enrolled at a single university. In contrast to the literature and perhaps surprisingly, we find no significant effect of the pandemic on both teaching quality and students' academic performance. This result holds along various dimensions such as family wealth, top-performance students, gender etc. Further, several robustness tests including running the estimation for courses not changed (neither lecturer nor term), matriculates only, mandatory courses and using multiple pre-treatment periods confirm the finding that the transition to online teaching due to the COVID-19 pandemic did not markedly affect SET and student performance.

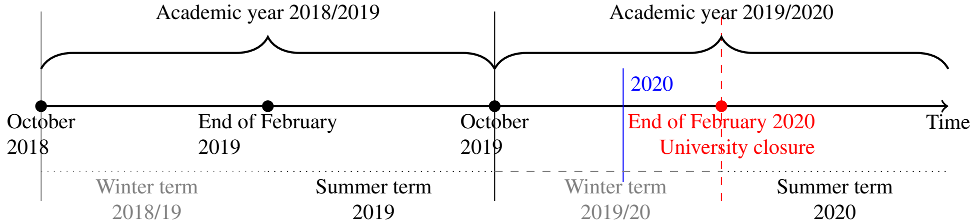
The remainder of this paper proceeds as follows. Section 2 describes the experimental setting and the data used. Next, Section 3 outlines the estimation approach and Section 4 presents the results. Section 5 looks a the effect of the COVID-19 pandemic on graduation grades and the percentage of students failing an exam, Section 6 conducts several robustness tests. Finally, Section 7 discusses the results and Section 8 concludes.

## 2 Experimental setting and data

In Italy, Lombardy was the region most hardly hit during the first wave of the pandemic in spring 2020. As a consequence, Lombardy was one of the first Italian regions to decide about school and university closures. Important for our experimental design is that the first university shutdown in Lombardy coincided with the beginning of the summer term 2020 at the University of Pavia. Closure of schools and universities in Lombardy, Piedmont, Veneto and Emilia-Romagna took place on February, 25 2020 and lasted until September 2020. At the University of Pavia, the start of lectures was scheduled on February, 24 2020 (and started then online one week later).

Figure 1 shows the timing of events. We start by considering the academic years 2018/2019 and 2019/2020, i.e. two time periods. Academic years are divided in two semesters or terms; the winter and summer term. The former generally goes from early October to late February, while the latter starts at the end of February and finishes in late September. We thus observe two winter terms and two summer terms. The closure of universities in Lombardy and the related online teaching in the summer term 2020 represents the treatment. The control group is represented by students evaluating and taking courses offered in the winter term, while the treatment group is defined by students evaluating and taking courses offered in the summer term.

Figure 1: Timeline of events




The University of Pavia is composed of 18 departments that offer 104 degree programs in total. We group six departments of medicine and engineering (surgery, internal, experimental and molecular medicine as well as civil and industrial engineering) in two departments (medicine and engineering). As these six departments are small ($<< 1,000$ observations each), grouping them in larger departments allows us to obtain more robust estimation results in case of separate regressions by departments. We, thus, remain with 14 different departments:

1. Natural science;
2. Chemistry;
3. Physics;
4. Law;
5. Engineering (composed of civil and industrial engineering);
6. Mathematics;
7. Medicine (composed of surgery, internal, experimental and molecular medicine);
8. Musicology;
9. Pharmacy;
10. Psychology;
11. Geology;
12. Economics and management;
13. Political and social science;
14. Humanities.

As stated before, we are interested in the effect of the COVID-19 pandemic along two dimensions: the effect on SET and the effect on students' performance (grades).[3] SET allow us to measure whether the level of teaching quality (from the student perspective) was kept during the pandemic, while student grades permit us to analyze the effect on student performance. Indeed, using student grades as measures of performance, we can answer the following questions. Has the pandemic had an effect on students' performance? Do we find negative effects for poorer students as in case of schools? That is, did family wealth play a role? Or, were there different effects for top- and bottom-performing students?

[3]Additionally, we look descriptively at the association of the pandemic on graduation grades and exam failure rates in Section 5.

SET were first introduced with the aim of providing feedback to lecturers about their teaching practices. Nowadays, SET are considered by deans and school managers as a tool to monitor 'customer satisfaction', and are often listed among the elements used to decide promotions and hiring. In particular, at the University of Pavia SET are used to grant biennial salary increases in lecturers' salaries. A national university reform has determined that faculty salary increases are no longer automatic but must meet criteria defined by the individual universities. According to our university's rules, every two years the lecturers may apply for a salary increase. The requirements that must be met are two out of three among the following: value of SET, number of publications produced, administrative commitment based on positions assigned in the university. Given that the number of administrative positions is limited, the condition on teaching evaluation becomes a necessary condition for salary increase. Since implementation of this rule for salary increase, violation of the criterion on SET was the only reason for not awarding the salary increase. Therefore, lecturers have – apart from an academic or personal interest – a strong monetary incentive to make a great effort for teaching.

We focus the analysis on bachelor and master courses, five-year degree in Law and six-year degree in Medicine. That is, we exclude degree programs no longer offered (*in disuso*).[4] We have two distinct data sets. The first data set consists of full record of SET (we proved details in Section 2.1). The second data set includes the universe of students that have taken at least one exam in the period considered. We have detailed information on each student such as number of exams taken, grades in each exam, date of exam, study degree, faculty, etc. as well as personal and family background information (details provided in Section 2.2).

## 2.1 Data on SET

We use administrative data including all SET produced by two cohorts of students in the academic years 2018/19 and 2019/2020. In order to register for an exam, students who attended the course must provide their evaluations and are redirected to the SET questionnaire. The completion of the SET questionnaire is a necessary condition for taking the exam. As compiling the questionnaire is fully anonymous, we cannot track the evaluations provided by a specific student.

Students are asked to evaluate the course by answering the following questions:

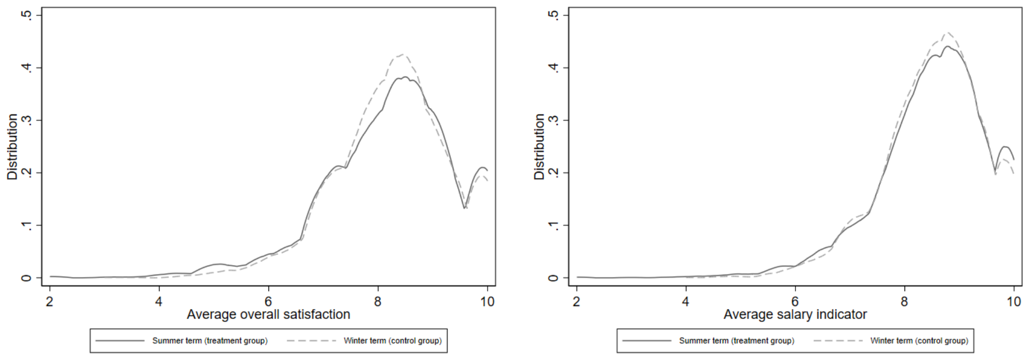
[4] In some departments, annual courses exist that we exclude from the analysis.

1. Overall satisfaction;
2. Lecturer's ability to motivate the class;
3. Lecturer teaches in clear way;
4. Lecturer is available for clarifications;
5. Clear presentation of learning objectives;
6. Quality of lecture notes/reference books;
7. Sufficient prerequisites;
8. Clear presentation of the exam rules from the beginning;
9. Lecturer present during office hours;
10. Workload is consistent with the ECTS;
11. Are the timetables for lectures, exercises and any other teaching activities respected?
12. Your interest for the subject.

The questionnaires include three further questions on tutorials. We do not consider them as they are not provided for all courses. Answers to SET are originally provided on a four-point Likert scale $\in [1, 4]$, where $1$ represents complete dissatisfaction with the course and $4$ complete satisfaction (see Lalla et al., 2005, for a discussion on the four-point Likert scale). The use of the four-point Likert scale was proposed by a research group appointed by the National Committee for University System Evaluation that also suggested to translate each category into the following values: $\{2, 5, 7, 10\}$, where $10$ represents complete satisfaction with the course and $2$ complete dissatisfaction. Following the literature (e.g. Bertoni et al., 2020) and the National Committee for University Evaluation (Valmon, 2020), we translate the scale into the values $\{2, 5, 7, 10\}$.

As most part of the literature (e.g. Bertoni et al., 2020), we take students' overall satisfaction (question 1 listed above) as the main indicator for SET. For details on the relation between overall satisfaction and the other indicators see Table A.1 in Appendix A. Additionally, we use the average of indicators 6, 7 and 8 that represents the indicator for salary increases. The criterion imposed by the university for salary increases stabilizes that the average of these three indicators needs to be at least equal to 7. Figure 2 presents the distribution of the two indicators (overall satisfaction and the average of the indicators for salary increase) per term. A two-sample Kolmogorov-Smirnov equality-of-distributions test shows that the difference in the indicators between the terms is at most 0.04 for overall satisfaction and 0.22 for the indicator relevant for salary increases. For the latter indicator, the difference is not statistically significant, while it is statistically significant at a 5% level for the former indicator.

Figure 2: Distribution of SET indicators by term




(a) Overall satisfaction

(b) Indicators for salary increase

Overall, we have 126,036 compiled questionnaires for the courses taught in the academic year 2018/2019 and 128,775 for those taught in the academic year 2019/2020 (see Table 1). Of these, 71,546 and 54,490 (74,450 and 54,325) questionnaires are for courses taught in the winter and in the summer term of the academic year 2018/19 (2019/20), respectively. In total, we observe 7,809 courses taught in these two academic years. We define a course as a learning unit taught by a specific lecturer in a specific academic year. Hence, we treat a given learning unit taught by the same lecturer in two academic years as two separate courses as the attendees belong to two different cohorts. In general, when several lecturers are involved in the learning unit, students fill a separate questionnaire for each lecturer.[5]

[5]There are lectures and tutorials, especially for matriculates, that are divided in separate courses. We treat these courses separately if they are taught by different lecturers, and as a single course if the same lecturer is in charge of all parts.

Table 1: Questionnaires compiled

| Academic year | Summer term | Winter term | Total |
|---|---|---|---|
| 2018/2019 | 54,490 | 71,546 | 126,036 |
| 2019/2020 | 54,325 | 74,450 | 128,775 |
| Total questionnaires | | 254,811 | |
| Total courses | | 7,809 | |
| Unique courses | | 2,117 | |

Table 2 shows descriptive statistics of SET by term for the full sample as well as separately by academic year. While both indicators of interest (overall satisfaction and relevant for salary increase) are higher in the summer compared to the winter term in academic year 2018/2019, the opposite holds in academic year 2019/2020. Note however, that the differences are quantitatively small despite being generally statistically significant. Differences in the number of questionnaires compiled and the fraction of women evaluating the course go in the same direction over the two academic years. Yet, on average students compile about two questionnaires less in the summer 2020 compared to the summer 2019. The fraction of women is stable over the years. All in all, the treatment (summer term) and control groups (winter term) are rather balanced across individual characteristics.

Table 2: Descriptive statistics SET by term, selected controls

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Summer term | | Winter term | | |
| Variable | Mean | Std.Dev. | Mean | Std.Dev. | Difference |
| *Panel (a) Full sample* | | | | | |
| Indicator for salary increase | 8.526 | 1.008 | 8.544 | 0.909 | -0.018 |
| Overall satisfaction | 8.239 | 1.185 | 8.285 | 1.034 | -0.046 |
| Number questionnaires compiled | 28.77 | 39.95 | 36.20 | 47.510 | -7.430*** |
| Women (fraction) | 0.611 | 0.278 | 0.603 | 0.263 | 0.008 |
| Observations | 3,779 | | 4,030 | | 7,809 |
| *Panel (b) 2018/2019* | | | | | |

*Continued on next page*

| Variable | Summer term Mean | Std.Dev. | Winter term Mean | Std.Dev. | Difference |
|---|---|---|---|---|---|
| Indicator for salary increase | 8.575 | 0.983 | 8.519 | 0.918 | 0.055* |
| Overall satisfaction | 8.272 | 1.145 | 8.263 | 1.041 | 0.009 |
| Number questionnaires compiled | 29.61 | 40.36 | 36.07 | 47.93 | -6.464*** |
| Women (fraction) | 0.600 | 0.272 | 0.596 | 0.257 | 0.003 |
| Observations | 1,838 | | 1,982 | | 3,820 |
| *Panel (c) 2019/2020* | | | | | |
| Indicator for salary increase | 8.480 | 1.030 | 8.569 | 0.899 | -0.089*** |
| Overall satisfaction | 8.207 | 1.220 | 8.307 | 1.026 | - 0.100*** |
| Number questionnaires compiled | 27.97 | 39.56 | 36.32 | 47.10 | -8.349*** |
| Women (fraction) | 0.621 | 0.284 | 0.609 | 0.268 | 0.011 |
| Observations | 1,941 | | 2,048 | | 3,989 |

*Notes:* Reported differences are based on a regression of the selected variables on a summer-term dummy. *, ** and *** denote significance at the 10%-, 5%- and 1%-level, respectively. Standard errors clustered at the course level are used.

## 2.2 University career data

We have data on the university career of all students enrolled at the university and having taken at least one exam in the period considered, i.e. either in academic year 2018/2019 or in academic year 2019/2020. For each student, we observe exam-specific grades, the overall average grade in all the exams taken, gender, date and place of birth, the municipality of residence, information about the degree course in which the student is enrolled and year of matriculation. Additionally, we have information about the enrollment status of students. That is, whether the student is a regular student (*studenti in corso*) or not. This classification distinguishes students on the basis of the length of their enrollment as compared to the official duration of the study programs. Additionally, we observe whether a course is mandatory or not.

We also observe each student's ISEE, which is an equalized economic situation indicator calculated on the basis of the family's yearly income and the family's non-labor income (e.g. assets). Further, the ISEE takes account of the family's composition (e.g. single parent, number and age of siblings). Tuition fees are paid based on the ISEE declaration submitted by students. The ISEE declaration is an official document issued by an official institution such as the municipality of residence. The University of Pavia defines tuition fees on the basis of 60 different income brackets based on the ISEE. Students who do not submit an ISEE declaration are assigned to the highest income bracket. Tuition fees vary from 0€ to 4,845€ per year and thus, in order to circumvent ending up in the highest bracket, students have a strong incentive to provide the ISEE declaration. As using the 60 income brackets defined by the university leads to few observations per bracket, we divide the ISEE in four different categories: bottom ($<$ 25th percentile), medium-bottom ($\geq$ 25th and $<$ 50th percentile), medium-top ($\geq$ 50th and $<$ 75th percentile), top ($\geq$ 75th percentile).

We exclude from the sample students older than 30 years. In order to form a more homogeneous sample of students. We calculate average values per student and term and we consider only exams that took place in the same term the course was taught. Finally, we end up with a sample of 149,376 exams for 18,415 unique students (see Table 3).

Table 3: Information on students and exams taken

| Academic Year | Winter term | Summer term | Total |
|---|---|---|---|
| 2018/19 | 25,177 | 42,749 | 67,926 |
| 2019/20 | 38,686 | 42,764 | 81,450 |
| Total exams | | 149,376 | |
| Total students | | 54,608 | |
| Unique students | | 18,415 | |

In Italy, university grades may vary $\in \{18, 31\}$, where $18$ represents the minimum grade for passing an exam and 31 represents the maximum grade (*30 e lode*). Students may reject the grade in case they are not satisfied with it and may repeat the exam as often as they want regardless whether they have passed it or not. In order to better understand whether the pandemic had an impact on the fraction of students failing an exam, we look in Section 5 at changes in failing an exam over time. Grades in Italian universities are generally not normally distributed (see Figure 3). A two-sample Kolmogorov-Smirnov equality-of-distributions test shows that the difference in average grades between the terms is at most 0.05 and that this difference is statistically significant at a 1% significance level. Figure A.1 in Appendix A shows the cumulative distribution function of average grades by term and academic year. The differences in distribution of average grades between the winter and summer term is not very pronounced. This finding holds for the full sample as well as for bachelor and master students.

Figure 3: Distribution of average grades by term

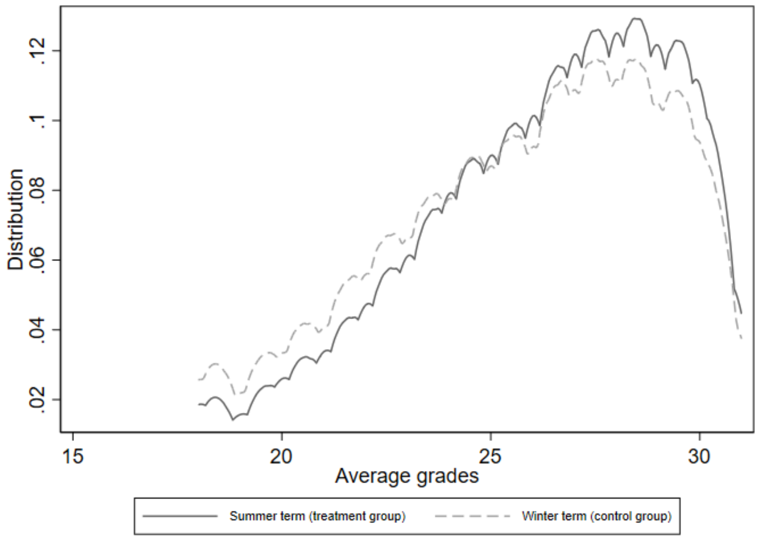



Table 4 shows descriptive statistics of students' university career data by term and academic year. On average students have statistically significantly higher average grades in the summer compared to the winter term. Similarly, students take significantly more exams in the summer compared to the winter. The data suggests only small differences in terms of ISEE, age and study type by term. Even though, the differences in summary statistics are generally statistically significant (presumably due to the relatively large number of observations), the magnitude of the differences is small. Thus, the treatment (summer term) and control groups (winter term) are well balanced across individual characteristics.

Table 4: Descriptive statistics student performance by term, selected controls

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Summer term | | Winter term | | |
| Variable | Mean | Std.Dev. | Mean | Std.Dev. | Difference |
| *Panel (a) Full sample* | | | | | |
| Average grade | 26.17 | 3.179 | 25.72 | 3.360 | 0.450*** |
| Age (years) | 23.03 | 2.210 | 23.09 | 2.223 | -0.060*** |
| Female (dummy) | 0.586 | 0.493 | 0.590 | 0.492 | -0.004*** |
| Regular student (dummy) | 0.858 | 0.349 | 0.853 | 0.354 | 0.005*** |
| ISEE bottom | 0.256 | 0.437 | 0.249 | 0.433 | 0.007*** |
| ISEE medium bottom | 0.249 | 0.433 | 0.248 | 0.432 | 0.001 |
| ISEE medium top | 0.233 | 0.423 | 0.235 | 0.424 | -0.002* |
| ISEE top | 0.261 | 0.439 | 0.268 | 0.443 | -0.010*** |
| Number of exams per term | 2.802 | 1.429 | 2.741 | 1.430 | 0.061*** |
| Bachelor (dummy) | 0.557 | 0.497 | 0.541 | 0.498 | 0.020*** |
| Master (dummy) | 0.192 | 0.394 | 0.196 | 0.397 | -0.004*** |
| Observations | 28,245 | | 26, 363 | | 54,608 |
| *Panel (b) 2018/2019* | | | | | |
| Average grade | 26.10 | 3.113 | 25.69 | 3.390 | 0.410*** |
| Age (years) | 23.35 | 2.109 | 23.36 | 2.115 | -0.010 |
| Female (dummy) | 0.584 | 0.493 | 0.592 | 0.492 | -0.008*** |
| Regular student (dummy) | 0.834 | 0.372 | 0.826 | 0.379 | 0.008*** |
| ISEE bottom | 0.248 | 0.432 | 0.246 | 0.431 | 0.002 |
| ISEE medium bottom | 0.249 | 0.432 | 0.246 | 0.431 | 0.003 |
| ISEE medium top | 0.239 | 0.426 | 0.239 | 0.426 | 0.000 |
| ISEE top | 0.265 | 0.441 | 0.269 | 0.443 | -0.005*** |
| Number of exams per term | 2.835 | 1.322 | 2.658 | 1.354 | 0.177*** |
| Bachelor (dummy) | 0.555 | 0.497 | 0.539 | 0.498 | 0.015*** |
| Master (dummy) | 0.176 | 0.381 | 0.175 | 0.380 | 0.001 |
| Observations | 11,771 | | 10,435 | | 22,206 |
| *Panel (c) 2019/2020* | | | | | |

*Continued on next page*

| Variable | Summer term Mean | Summer term Std.Dev. | Winter term Mean | Winter term Std.Dev. | Difference |
|---|---|---|---|---|---|
| Average grade | 26.22 | 3.224 | 25.74 | 3.340 | 0.480*** |
| Age (years) | 22.79 | 2.250 | 22.92 | 2.275 | -0.130*** |
| Female (dummy) | 0.587 | 0.492 | 0.588 | 0.492 | -0.001 |
| Regular student (dummy) | 0.875 | 0.331 | 0.871 | 0.336 | 0.004 |
| ISEE bottom | 0.262 | 0.440 | 0.251 | 0.434 | 0.011*** |
| ISEE medium bottom | 0.250 | 0.433 | 0.249 | 0.432 | 0.001 |
| ISEE medium top | 0.230 | 0.421 | 0.233 | 0.423 | -0.003*** |
| ISEE top | 0.259 | 0.437 | 0.267 | 0.442 | -0.002 |
| Number of exams per term | 2.778 | 1.501 | 2.796 | 1.476 | -0.018 |
| Bachelor (dummy) | 0.559 | 0.497 | 0.573 | 0.498 | 0.004** |
| Master (dummy) | 0.203 | 0.402 | 0.210 | 0.407 | -0.007*** |
| Observations | 16,474 | | 15,928 | | 32,402 |

*Notes:* Reported differences are based on a regression of the selected variables on a summer-term dummy. *, ** and *** denote significance at the 10%-, 5%- and 1%-level, respectively. Standard errors clustered at the individual level are used.

# 3 Estimation approach

Our empirical framework is based on a difference-in-differences (DID) approach. We start by considering the simple two-group two-period DID model. For robustness, we extend the setup by considering multiple pre-treatment periods in Section 6. This extension allows us using and testing a more flexible common trend assumption. Under the common trend assumption, the relevant unmeasured variables are either time-invariant group attributes or time-varying factors that are group invariant. Together, these restrictions imply that the time series of outcomes in each group should differ by a fixed amount in every period and should exhibit a common set of period-specific changes (see Section 6 for details). Further, in our setup, we can clearly assume strict exogeneity. The timing of treatment exposures in our DID design is statistically independent of the potential outcome distributions conditional on the group- and time-fixed effects.

We use the DID strategy to assess the effects of the transition to online teaching due to the COVID-19 pandemic on higher education along two dimensions: educational quality and student performance. In case of educational quality, we analyze the impact of the COVID-19 pandemic on SET completed by students. In case of student performance, we evaluate the effects of the pandemic on students' average grades. Consequently, we apply two distinct identification strategies of the DID (one for SET indicators and one for students' average grades) in order to analyze both sides of the teaching process.

## 3.1 Estimation for SET

We have two groups defined by the term when the course was taught and two time periods defined by the academic years 2018/2019 and 2019/2020. The control group is given by SET of courses taught in the winter term, while the treatment group is composed of SET of courses taught in the summer term. In academic year 2019/2020, only the treatment group is affected by the treatment, i.e. by online teaching that was unexpectedly imposed on all courses.

To evaluate the treatment effect of the unexpected transition from presence to online teaching on SET, we run two linear regressions according to the following equation:

$$y^k_{jst} = \alpha^k + \gamma^k d_s + \delta^k post_t + \beta^k d_s * post_t + u^k_{jst} \qquad \text{with } k = 1{,}2 \tag{1}$$

where $y_{jst}$ is the average overall-satisfaction ($k = 1$) or salary-relevant indicator ($k = 2$) obtained by all students in a specific course $j$ in term $s$ and year $t$. $d$ is the dummy variable indicating the group status: $d = 1$ when the exam is taken in the summer and $d = 0$ when the exam is taken in the winter term. $post$ is a binary variable taking value $0$ in the baseline (academic year 2018/19) and value $1$ in the follow-up year (academic year 2019/2020) and $u_{jst}$ is the corresponding error term. The interaction term $d * post$ defines treatment: the summer term of academic 2019/2020. That is, when the pandemic hit the university.

We consider also a specification that contains a vector of controls $z_{jst}$. The latter includes control variables such as the number of questionnaires compiled per course, fraction of women per course and lecturer fixed effects. In a given time interval $st$, we observe $J$ courses and $L$ lecturers. The function $L(j, s, t)$ identifies the unique lecturer that is evaluated in course $j$ in term $s$ and academic year $t$ (i.e., in the time interval $st$). In case of the alternative specification, the estimation equation reads then as:

$$y^{k}_{jst} = \alpha^{k,full} + \gamma^{k,full} d_s + \delta^{k,full} post_t + \beta^{k,full} d_s * post_t + z_{jst}\lambda^{k,full} + \sum_{l=1}^{L} \phi_l^{k,full} * h_l + u_{jst}^{k,full} \quad (2)$$

where $h_l$ is a dummy for lecturer $l$ that is evaluated in course $j$ in term $s$ and academic year $t$, with $l = L(j, s, t)$. Observe that *full* identifies the coefficient estimates of the alternative specification.

If students are randomly distributed across courses, or if all courses must be attended by all students, then the distribution of students' reporting styles is the same in all courses and reporting heterogeneity does not bias the relative evaluation of a course. In order to account for this bias, we restrict the sample to mandatory courses and repeat the analysis for this subsample in Section 6.

## 3.2 Estimation for student performance

In order to evaluate the impact of the treatment (transition to online teaching due to the legally imposed university closure) on student performance, we compare average grades in the summer and winter term over the academic years before (2018/2019) and after (2019/2020) its implementation. This procedure allows us to identify – other things equal – the causal effect of online teaching on student performance.

In our design, the control group consists of students having attended courses taught in the winter, while the treatment group refers to students having attended courses taught in the summer term. We estimate the following equation:

$$v_{ist} = \alpha + \gamma d_s + \delta post_t + \beta d_s * post_t + \epsilon_{ist} \quad (3)$$

where $v_{ist}$ is the average grade obtained by student $i$ in the exams of term $s$ and year $t$ and $\epsilon$ is the corresponding error term. As in case of the SET in equation (1), $d$ represents a dummy variable for the summer term and $post$ is a dummy for academic year 2019/2020. The interaction term $d * post$ defines again the treatment or the summer term 2020.

In an alternative specification, we add a vector of controls $\mathbf{x}_{ist}$ that accounts for age, gender, being a regular student, year of matriculation, ISEE, number of exams taken and study program. Further, in this alternative specification, we use as dependent variable the average grade depurated from lecturer fixed effects. The correction or depuration proceeds in two steps.[6] First, we estimate lecturer fixed effects in the following regression: $v_{clst} = \alpha_0 + c_c * \iota_c + \theta_l + \phi_s + \omega_t + \tilde{\epsilon}_{lcst}$, where $\theta_l$, $\phi_s$ and $\omega_t$ represent lecturer, term and academic year fixed effects, respectively. $c$ are course or exam dummies, $\alpha_0$ is an intercept and $\tilde{\epsilon}$ is the corresponding error term. Second, the corrected grades net of the lecturer time-constant heterogeneity or the lecturer fixed effects are defined as: $\tilde{v}_{iclst} = v_{iclst} - \hat{\theta}_l$. Observe that student $i$ writes exam $c$ of lecturer $l$ in term $s$ and academic year $t$. The corrected grades are then averaged over exams $c$ of student $i$ in term $s$ and academic year $t$: $\bar{\tilde{v}}_{ist} = \tilde{v}_{ist}$. The latter is the dependent variable in the alternative specification.

As courses may be relocated to a different term or the lecturer may change over time, we restrict in Section 6 the analysis to courses that did not change across these two dimensions. In order to consider two groups of students as homogeneous as possible, we also look at the impact of the transition to online teaching due to the pandemic on matriculates and mandatory courses (see Section 6).

# 4 Empirical results

In this section, we present the estimation outcome. We start by providing the results for SET and then proceed to the estimation outcome for student performance. We conduct the analysis for SET for the full sample as well as separately by department. In case of student performance, we consider additionally various sociodemographic dimensions such as ISEE category or gender.

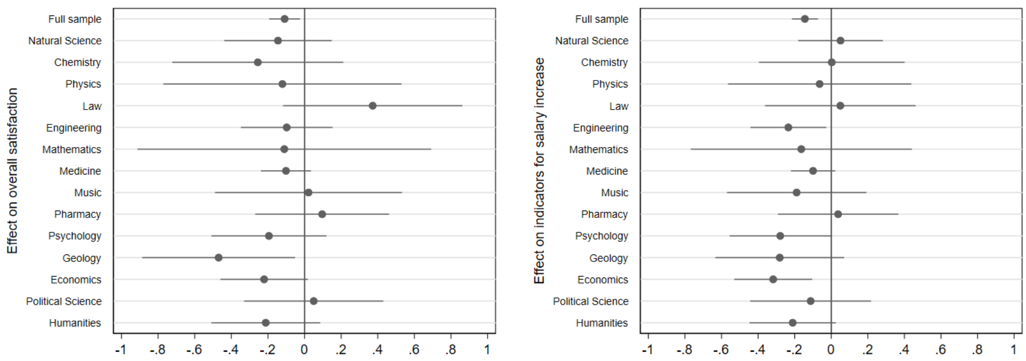
## 4.1 SET

As stated above, we present the effects on the indicator of overall satisfaction and on the indicator used by the university to decide about salary increases. Figure 4 shows the aggregate effects (i.e. specification without control variables) of the COVID-19 pandemic on SET for the full sample as well as separately by department. In the full sample of all courses offered by the university, we see a statistically significant reduction in the satisfaction levels of both indicators. Yet, this effect is small and

[6] Observe that the correction is similar to the approach of Canay (2011) for panel data (see e.g. Bargain et al., 2018; Bonaccolto-Töpfer et al., 2021; Castagnetti and Giorgetti, 2019, for empirical applications).

mostly no longer statistically significant when analyzing the departments separately – presumably due to less numbers of observations per department (see Table A.2 for number of observations). Except for the departments of geology, engineering and economics, we never find a statistically significant effect on SET. That is, in only one-fifth of the departments, we find a statistically significant effect at all. In all cases, the effects are small, amounting to -0.5 in the department of geology at most. The latter represents a reduction in overall satisfaction of 6% (given an average overall satisfaction indicator of 8.3 in this department) due to the first wave of the COVID-19 pandemic.

Figure 4: Effect on course SET of online teaching due to the COVID-19 pandemic, full sample (no controls added)

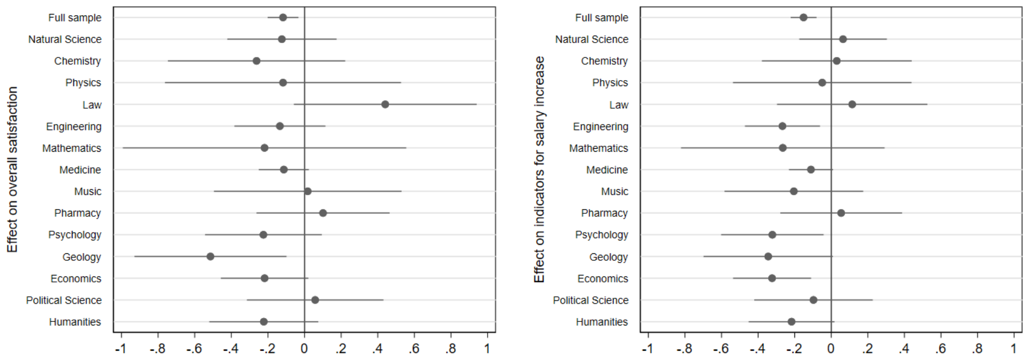



(a) Indicator for overall satisfaction (b) Indicator for salary increase

*Notes:* Estimates of effects on SET for the full sample and by departments. The figure shows estimates of the effects on SET from a difference-in-differences specification that compares values on SET for the courses taught in the winter and summer term. Point estimates with 95% confidence intervals calculated based on standard errors clustered at the course level. Table A.2 shows the corresponding number of observations.

The results from the alternative specification show that there is also only a slight negative effect on SET due to the pandemic when controlling for general observable characteristics (see Figure 5). Even though the effect for the full sample is again statistically significantly different from zero, it is quantitatively negligible. To be precise, we find that students evaluate the courses about 0.2 points or 2% worse in the full sample. This finding holds for both indicators. The effect on overall satisfaction across different departments is only statistically significant for geology. In case of the indicators for salary increase, we find statistically significant, though small, effects for the department of engineering, psychology and economics. That is, we find heterogeneous effects across the distinct departments and the different indicators. Yet, quantitatively the effects are small throughout.If statistically significant, the effects are always negative. The latter implies that, if at all, teaching quality is affected negatively. However, as the effects are at most modest, we conclude that the COVID-19 pandemic did not have a relevant effect on SET.

Figure 5: Effect on course SET of online teaching due to the COVID-19 pandemic, full sample (controls added)




(a) Indicator for overall satisfaction (b) Indicator for salary increase

*Notes:* Estimates of effects on SET for the full sample and by departments. The figure shows estimates of the effects on SET from a difference-in-differences specification that compares values on SET for the courses taught in the winter and in the summer term. Control variables used are number of questionnaires compiled, female share, course year and lecturer fixed effects. Point estimates with 95% confidence intervals calculated based on standard errors clustered at the course level. Table A.2 shows the corresponding number of observations.

Overall, except for few departments (engineering, psychology, economics and geology), we never find statistically significant effects of the COVID-19 pandemic on SET. Also, in case of statistically significant effects, they are negligibly small. This finding holds for the specification with and without control variables. As we consider SET as a proxy for teaching quality, we do not find an effect of the pandemic on higher educational quality. The results add to the COVID-19 literature suggesting that students did not evaluate courses substantially different in the first wave of the pandemic. Hence, our results do not sustain concerns of a reduction in educational quality due to the pandemic. We discuss potential implications of this finding in Section 7.

## 4.2 Students' performance

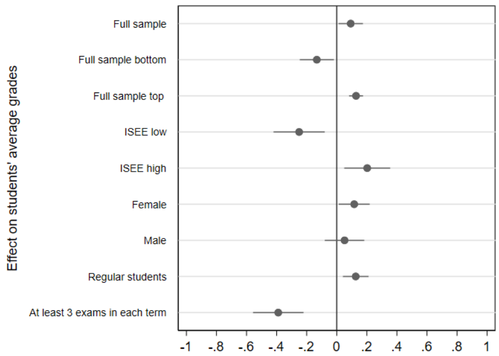
Here, we investigate the effect of the COVID-19 pandemic on student performance. Apart from looking at the full sample, bachelor and master students and different departments, we inter alia look separately at top and bottom students as well as at the role of family background for student performance. Further, in order to rule out gender effects, we consider male and female students separately. Similarly, we run the analysis for regular students and students taking at least three exams per term.

We define as top (bottom) students those students whose average grades are above or equal to the 75th (below the 25th) percentile of the distribution of grades before the summer term 2020. We identify students from richer or poorer families via the ISEE. Students with low ISEE report an ISEE below the 25th percentile of the income brackets defined by the university, while students with high ISEE report an ISEE equal to or above the 75th percentile of the income brackets defined by the university.

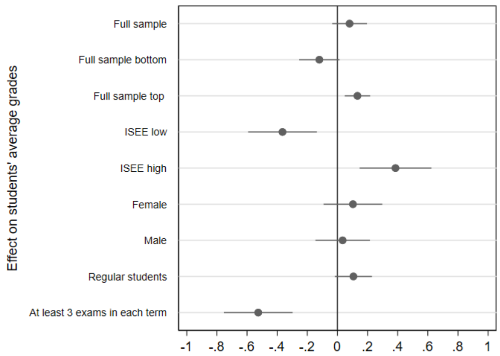
Figure 6 shows the aggregate effect of the COVID-19 pandemic on student performance for the full sample as well as for selected subsamples (panel (a)). The aggregate effect comes from the specification without control variables. We find a slightly positive, statistically significantly, though negligibly small effect for the full sample (amounting to 0.08 grade points). Bottom students are affected negatively, while top students are affected positively. If verified, this finding would suggest that bottom students fall behind due to the pandemic. However, the effects in both directions are small. Further, the results suggest negative effects for students from poorer but positive effects for students from richer families. The latter would imply that family wealth has played a role and that the pandemic has increased educational inequality along this socioeconomic dimension. For men and women, we do not find statistically significantly different effects (overlapping confidence intervals). The effects are slightly positive for women, while we find no effect for men. The transition to online teaching has affected regular students' average grades slightly positive. Students taking three or more exams per term suffer in terms of average grades from the COVID-19 pandemic. However, the effect size, even though generally statistically significant, is small amounting to 0.43 grade points for students taking at least three exams in the full sample. The latter amounts to a decline in average grades (26.5) of 1.6% for this subsample.

The pattern of these results persists generally also for bachelor and master students (panel (b) and (c), respectively). Looking separately at bachelor and master students suggests that the negative effect on performance for students from poorer families or students taking at least three exams is driven by bachelor students, while the slightly more positive effect for women is driven by master students.

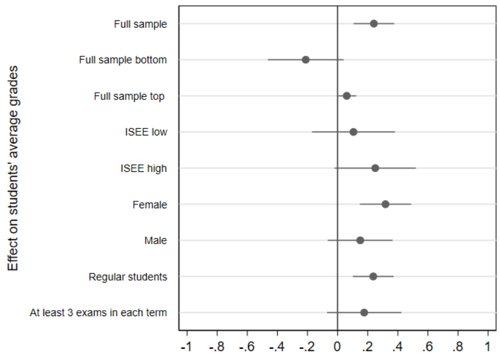
Figure 6: Effects on students' average grades of online teaching due to the COVID-19 pandemic (no controls added)




(a) Full sample




(b) Bachelor




(c) Master

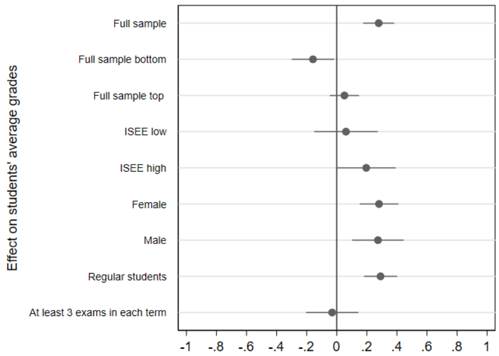
*Notes:* The figure shows estimates of the effects on students' average grades from a difference-in-differences specification without controls. Point estimates with 95% confidence intervals, standard errors clustered at the individual level. Table A.3 shows the corresponding number of observations.

Figure 7 shows the results from the alternative specification with control variables for the full sample as well as for selected subsamples. In this specification, the effect size for the full sample is slightly higher compared to the specification without control variables amounting to 0.22 grade points. Yet, again, this effect is quantitatively small representing an increase in average grades (26) due to the pandemic of 0.8%. The results suggest still slightly statistically significantly negative effects for bottom students, while they suggest no effect for top students. We no longer find statistically significantly effects for both students from poorer and richer families. Moreover, as the confidence bands do overlap in case of bottom and top students as well as in case of students from richer and poorer families, the corresponding point estimates do no longer statistically significantly differ. Thus, the pattern of adverse effects for bottom and top students or students from richer and poorer families disappears. The effects for both men and women are positive, though small and do again not differ statistically significantly from each other. Regular students tend to gain from the pandemic in terms of average grades. However, the effect is again small. In case of the specification with control variables, students doing at least three exams do no longer suffer from the COVID-19 pandemic in terms of grades. The findings for the full sample do not change substantially for bachelor and master students.

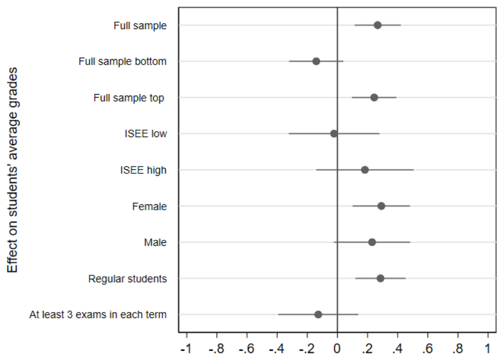
Overall, the results suggest no substantial effect of the first wave of the COVID-19 pandemic on average grades. The general tendency of average grades is that they increased due to the transition to online teaching. Yet, the effects are tiny throughout (amounting generally to less than 1%). We discuss potential implications of these results in Section 7. Our findings are robust along various socioeconomic dimensions and, thus, stand in contrast to the results for schools. Indeed, the literature on effects of the COVID-19 pandemic for schools found substantial differences in effects for men and women (Engzell et al., 2020), for top and bottom students (Aucejo et al., 2020) or for students from poorer and richer families (Agostinelli et al., 2020; Bacher-Hicks et al., 2021; Chetty et al., 2020).

Figure 8 shows the effect of the COVID-19 pandemic on student performance by departments conditional on control variables. Figure C.3 in Appendix C presents the corresponding figure without controls. The effects remain small across all departments and for both bachelor (panel (b)) and master (panel (c)) students. Nonetheless, the results suggest heterogeneity across departments such as positive effects on grades of psychology students. We find no systematic pattern for bottom or top performing students.

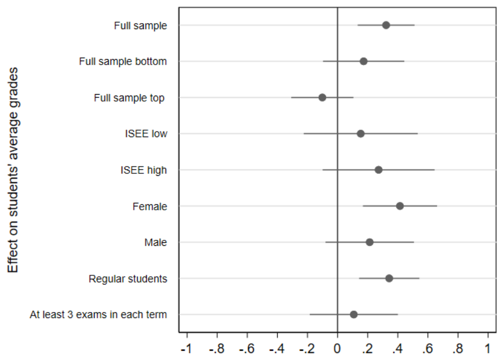
Figure 7: Effects on students' average grades of online teaching due to the COVID-19 pandemic (controls added)




(a) Full sample



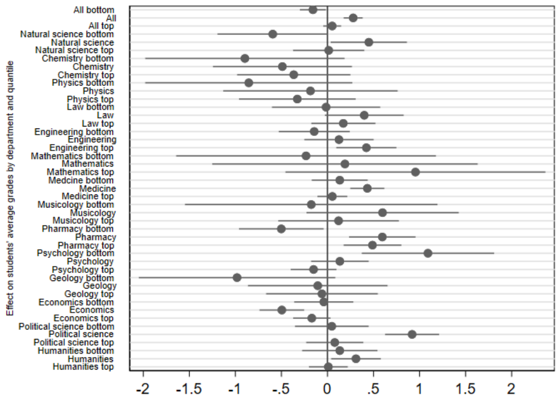

(b) Bachelor



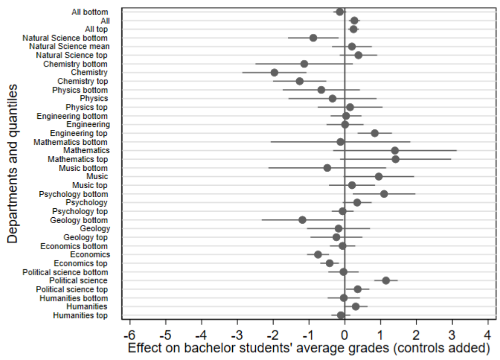

(c) Master

*Notes:* The figure shows estimates of the effects on students' average grade from a difference-in-differences specification with controls. Control variables used are gender, age, dummies for being a regular student, ISEE, master, 6-year degree or 5-year degree and year of matriculation. The dependent variable is the student-level average of corrected grades $\tilde{v}_{ist}$. Standard errors clustered at the individual level are used. Point estimates with 95% confidence intervals. Table A.3 shows the corresponding number of observations.

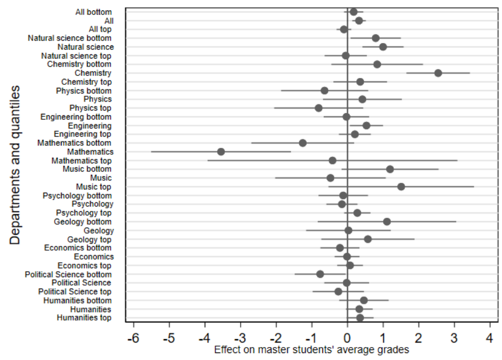
Figure 8: Effects on students' average grades of online teaching due to the COVID-19 pandemic by departments and for top and bottom students (controls added)




(a) Full sample




(b) Bachelor




(c) Master

*Notes:* The figure shows estimates of the effects on students' average grade from a difference-in-differences specification with controls. Control variables used are gender, age, dummies for being a regular student, ISEE, master, 6-year degree or 5-year degree and year of matriculation. The dependent variable is the student-level average of corrected grades $\tilde{v}_{ist}$. Top students are defined as students with average grades above or equal to the 75th percentile of the grade distribution before treatment. Bottom students are students with average grades below the 25th percentile of the grade distribution before treatment. Standard errors clustered at the individual level are used. Point estimates with 95% confidence intervals. Table A.3 shows the corresponding number of observations.

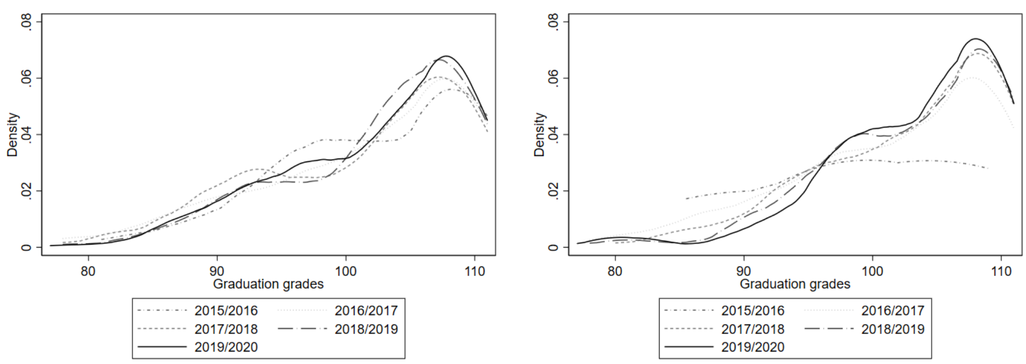
To sum up, we find modest to no effects of the transition from presence to online teaching in the summer term 2019/2020 on students' average grades. This finding holds for various subsamples as well as for bachelor and master students. With regard to different departments, we find heterogeneous, though small, effects throughout. As stated, in contrast to the recent literature on the effects of the COVID-19 pandemic on schools, we find no indication of adverse effects on students performance for top or bottom students or for students from poorer and richer families. Similarly, we find no pronounced gender differences in effects.

# 5 Changes in graduation and failing an exam over time

In this section, we look descriptively at the impact of the COVID-19 pandemic on graduation grades and failures of exams. We have aggregate data on graduation grades and exam failures, withdrawals or non-acceptance between the academic years 2015/2016 and 2019/2020. Recall that students may reject grades and may repeat the exam as often as they want regardless whether they have passed it or not. As we cannot disentangle the number of students that failed exams from those that withdrew or did not accept the grade, we rely on a proxy of exam failures consisting of all three potential reasons.

Figure 9 shows the density of average graduation grades over time. We find no substantial change in the densities of graduation grades over time and in particularly not in the summer term 2019/2020, i.e. the semester COVID-19 hit the university. Thus, we conclude that graduation both in terms of number of graduates and grades at this university was not affected by the pandemic. This finding stands in contrast to results from the literature on the effects of economic recessions on graduation (e.g. Oreopoulos et al., 2012).

Figure 9: Density of graduation grades over time

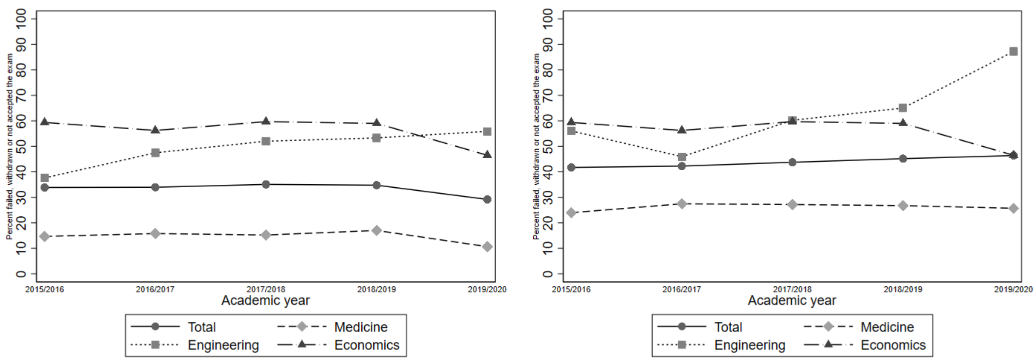



(a) Summer term (b) Winter term

*Notes:* In Italy, students may graduate with a final grade between 72-111, where 111 represents *110 e lode*. In the winter term 2015/2016 substantially less graduation dates were offered.

Figure 10 shows the percentage of exams failed over time for the full sample as well as for the three largest departments (medicine, engineering and economics) according to our data. The data suggests no significant effect of the COVID-19 pandemic on this measure. Engineering is the only department, where we observe an increase over time. This increase is considerable in the winter but small in the summer term 2019/2020. Consequently, the increase is not attributable to the COVID-19 pandemic. For the full sample as well as the remaining departments (medicine and economics), we find a slightly negative effect in the summer term 2020. The decline, however, is small amounting to two percentage points. Hence, we conclude that the number of exam failures was at most slightly affected by the pandemic and that this effect was generally negative. Our data does thus not support concerns of more students falling behind due the first wave of the pandemic.

Figure 10: Percentage of exams failed over time, full sample and selected departments




(a) Summer term (b) Winter term

*Notes:* 'Exams failed' includes students that failed the exam, withdraw from an exam or did not accept the grade of the exam. That is, it represents a proxy for exam failures. The percentage of 'exams failed' is calculated as: $\frac{\text{Number of exams failed, withdrawn or not accepted}}{\text{Number of total exams passed}} * 100$ per term.

# 6 Robustness tests

In this section, we repeat the main analysis for different subsamples. First, we restrict the analysis to courses where neither the term when course took place nor the lecturer has changed. Second, we focus on matriculates only. Third, we look at mandatory courses only. Finally, we use multiple pre-treatment periods starting in academic year 2015/2016. For simplicity, we focus here on the specification with control variables. However, we provide the estimation outcome without control variables in Appendix C. The main insights do not change.

## 6.1 Courses not changed

We repeat the analysis on the subsample of courses not changed. That is, neither the term when the course took place (summer or winter) nor the lecturer has changed between academic years 2018/2019 and 2019/2020. As we have again two periods and two groups in this robustness test, we can maintain the experimental setting (Section 2) as well as the estimation approach (Section 3).

**Results SET**

Table 5 shows descriptive statistics for SET in case of courses not changed. The sample restriction to courses not changed leaves us with 6,574 observations (compared to 7,809 in the main analysis). Indicators relevant for salary increase as well as overall satisfaction remain stable to this restrictions (in case of the full sample, we had: 8.53 and 8.24 for the summer and 8.54 and 8.30 for the winter term, respectively). Also the fraction of women evaluating the course did not change (in case of the full sample, we had: 0.61 and 0.60 in the summer and winter, respectively). The same holds for the number of questionnaires compiled.

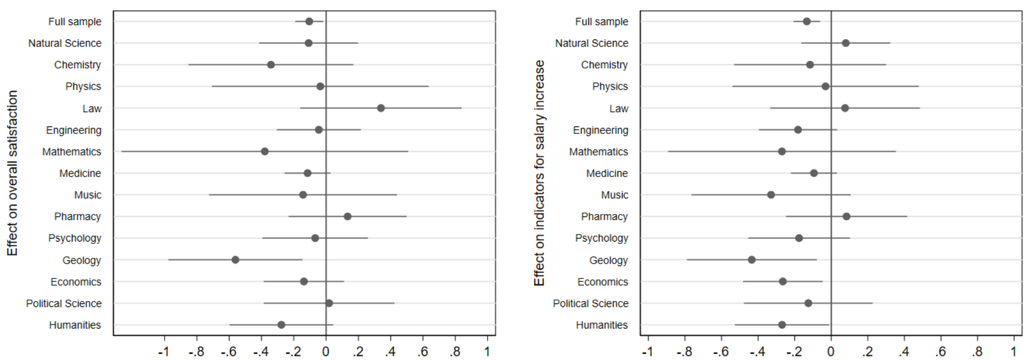
Table 5: Descriptive statistics SET by term for courses not changed, selected controls

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Summer term | | Winter term | | |
| Variable | Mean | Std.Dev. | Mean | Std.Dev. | Difference |
| Indicator for salary increase | 8.497 | 0.998 | 8.522 | 0.897 | -0.025 |
| Overall satisfaction | 8.218 | 1.174 | 8.268 | 1.015 | -0.050 |
| Number questionnaires compiled | 30.480 | 40.860 | 37.750 | 48.880 | -7.270*** |
| Women (fraction) | 0.611 | 0.273 | 0.605 | 0.261 | 0.006 |
| Observations | 3,080 | | 3,494 | | 6,574 |

*Notes:* Reported differences are based on a regression of the selected variables on a summer-term dummy. *, and denote significance at the 10%-, 5%- and 1%-level, respectively. Standard errors clustered at the course level are used.

Figure 11 shows the conditional effects of the COVID-19 pandemic on SET for the subsample of courses not changed. As in case of the main analysis, we find a statistically significant and negative effect for the full sample for both indicators. Yet, the effects are again negligibly small. The effects turn statistically insignificant for all departments except for geology in case of the overall-satisfaction indicator and for geology, economics and humanities in case of the indicator for salary increases. Again, the overall effects are tiny. The effect is largest for the overall-satisfaction indicator in the department of geology amounting to 0.59 points (or 7% given an average indicator of 8.23). The effects are thus rather small and comparable to the results from the main analysis. Further, only three out of 14 departments are affected statistically significantly. Hence, educational quality at the University of Pavia is not at stake.

Figure 11: Effect on course SET of online teaching due to the COVID-19 pandemic, courses not changed (controls added)




(a) Indicator for overall satisfaction (b) Indicator for salary increase

*Notes:* Estimates of effects on SET for courses not changed and by departments. Full sample refers to the sample of courses not changed. The figure shows estimates of the effects on SET from a difference-in-differences specification that compares values on SET for the courses taught in the winter and in the summer term. Control variables used are number of questionnaires compiled, female share, course year and lecturer fixed effects. Point estimates with 95% confidence intervals calculated based on standard errors clustered at the course level. Table A.2 shows the corresponding number of observations.

### Results student performance

Table 6 shows descriptive statistics for student performance for the subsample of courses not changed. In total, we remain with 49,609 observations (compared to

54,608 in the main analysis). Average grades are again slightly higher in the summer than in the winter term, while the fraction of women and students' age are rather constant over the academic years. The same holds for the portion of regular students, ISEE indicators and bachelor and master students. Students take on average 2.8 exams in the summer and 2.7 exams in the winter. Thus, the difference over the terms is comparable to that of the main analysis (0.06).

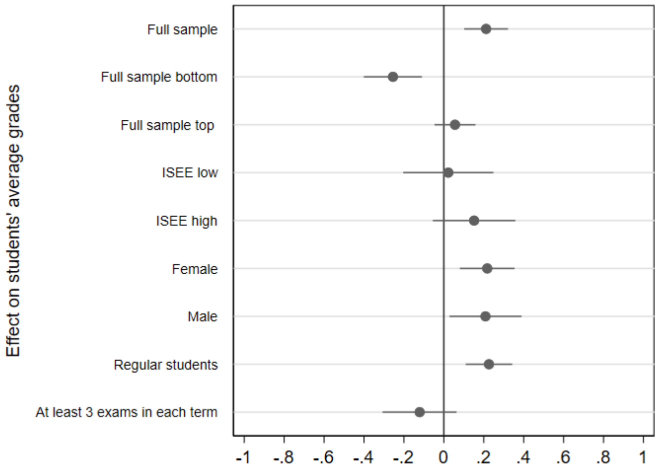
Table 6: Descriptive statistics student performance for courses not changed, selected controls

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Summer term | | Winter term | | |
| Variable | Mean | Std.Dev. | Mean | Std.Dev. | Difference |
| Average grade | 26.08 | 3.191 | 25.66 | 3.379 | 0.420*** |
| Age (years) | 22.97 | 2.202 | 23.06 | 2.219 | -0.09*** |
| Female (dummy) | 0.584 | 0.493 | 0.589 | 0.492 | -0.005** |
| Regular student (dummy) | 0.856 | 0.351 | 0.853 | 0.354 | 0.003 |
| ISEE bottom | 0.258 | 0.437 | 0.247 | 0.431 | 0.011*** |
| ISEE medium bottom | 0.248 | 0.432 | 0.247 | 0.431 | 0.001 |
| ISEE medium top | 0.232 | 0.422 | 0.236 | 0.424 | -0.004** |
| ISEE top | 0.262 | 0.439 | 0.269 | 0.443 | 0.007*** |
| Number of exams per term | 2.792 | 1.433 | 2.733 | 1.430 | 0.059*** |
| Bachelor (dummy) | 0.565 | 0.496 | 0.544 | 0.498 | 0.019*** |
| Master (dummy) | 0.175 | 0.380 | 0.185 | 0.388 | -0.01*** |
| Observations | 25,397 | | 24,212 | | 49,609 |

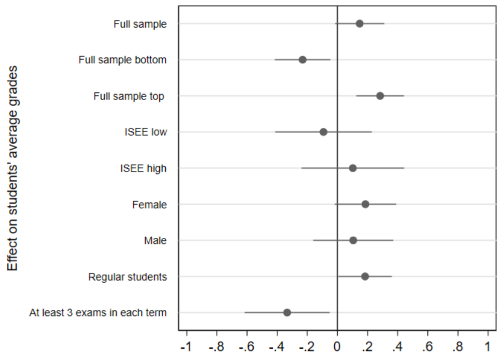
*Notes:* Reported differences are based on a regression of the selected variables on a summer-term dummy. *, and denote significance at the 10%-, 5%- and 1%-level, respectively. Standard errors clustered at the individual level are used.

Figure 12 presents the estimation outcome for selected subgroups. As in the main analysis, we find only small to no effects of the COVID-19 pandemic on average grades. For the full sample of courses not changed, the effect equals to 0.31 grade points or 1.2% (given average grades of 25.9 in the sample of courses not changed). The results suggest no statistically significant effect for top students, but negative effects for bottom students. These effects are driven by bachelor students (panel (b)). Both students from poorer and richer families are not affected in terms of grades by the pandemic. Further, we do not find statistically significantly different effects for students from richer or poorer families as well as for men or women (overlapping confidence bands, respectively). In case of courses not changed, only bachelor students taking at least three exams perform statistically significantly worse due to the pandemic (panel (a)). All in all, the effects are again either statistically insignificant or negligibly small.

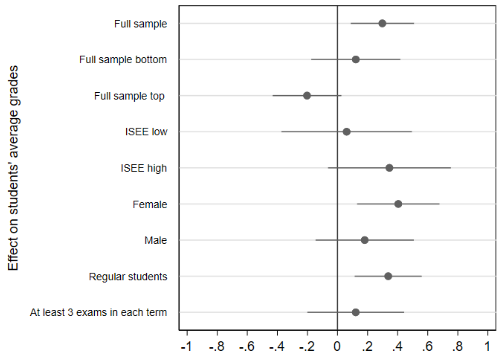
Figure 12: Effects on average grades of online teaching due to COVID-19 pandemic, courses not changed (controls added)




(a) Full sample




(b) Bachelor




(c) Master

*Notes:* Estimates of effects on average grades. Full sample refers to courses not changed. The figure shows estimates of the effects on students' average grade from a difference-in-differences specification with controls. Control variables used are gender, age, dummies for being a regular student, ISEE, master, 6-year degree or 5-year degree and year of matriculation. The dependent variable is the student-level average of corrected grades $\tilde{v}_{ist}$. Standard errors clustered at the individual level are used. Point estimates with 95% confidence intervals. Table A.3 shows the corresponding number of observations.

## 6.2 Matriculates only

Here we consider only first-year students (matriculates). This subsample represents a more homogeneous group compared to considering all students. Students may enroll both in the winter and the summer term at the University of Pavia. Thus, with the aim of generating a homogeneous sample, we restrict the sample to students having enrolled in the academic years 2018/2019 and 2019/2020. As a consequence of this sample restriction, we consider the impact of the pandemic on SET and average grades for the courses offered to first-year students only.

The experimental setting remains unchanged. Again, we have two periods and two groups. The cohort of students matriculated in the academic year 2018/2019 represents the control group. The cohort of students that has matriculated in the academic year 2019/2020 is the treatment group. Treatment occurs in the summer term of the academic year 2019/2020, i.e. when the pandemic hit the university.

**Results SET**

Table 7 shows descriptive statistics for SET of matriculates. The SET indicators are slightly lower in the summer compared to the winter. The corresponding differences are statistically insignificant. On average, matriculates compile eight questionnaires less in the summer term. We observe equal fractions of female matriculates over the academic years. Overall, the statistics are similar to those from the main analysis.

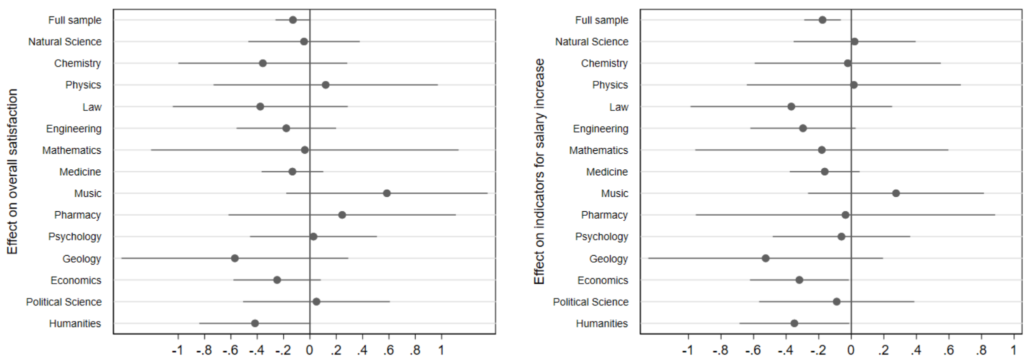
Table 7: Descriptive statistics SET by semester matriculates, selected controls

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Summer term | | Winter term | | |
| Variable | Mean | Std.Dev. | Mean | Std.Dev. | Difference |
| Indicator for salary increase | 8.505 | 1.048 | 8.560 | 0.933 | -0.050 |
| Overall satisfaction | 8.178 | 1.217 | 8.306 | 1.061 | -0.130 |
| Number questionnaires compiled | 27.606 | 39.616 | 35.236 | 47.501 | -7.630*** |
| Women (fraction) | 0.601 | 0.289 | 0.589 | 0.273 | 0.003 |
| Observations | 1,688 | | 1,609 | | 3,397 |

*Notes:* Reported differences are based on a regression of the selected variables on a summer-term dummy. *, and denote significance at the 10%-, 5%- and 1%-level, respectively. Standard errors clustered at the course level are used.

Figure 13 shows the effects on SET of the COVID-19 pandemic for matriculates. As in the previous cases, we find no pronounced effects. Moreover, effect heterogeneity across departments is reduced. In case of the overall satisfaction indicator, we find no statistically significant effect at all. In case of the indicator for salary increase, only SET in the full sample and the economics department are statistically significantly lower. Overall, we find at most in one out of fourteen departments a statistically significantly effect of the transition to online teaching due to the COVID-19 pandemic in spring 2020 on SET. That is, also for the homogeneous subsample of matriculates, SET are not markedly affected by the transition to online teaching during the summer term 2020.

Figure 13: Effect on course SET of online teaching due to the COVID-19 pandemic, matriculates (controls added)




(a) Indicator for overall satisfaction (b) Indicator for salary increase

*Notes:* Estimates of effects on SET for matriculated and by departments. Full sample refers to the sample of matriculates. The figure shows estimates of the effects on SET from a difference-in-differences specification that compares values on SET for the courses taught in the winter and in the summer term. Control variables used are number of questionnaires compiled, female share, course year and lecturer fixed effects. Point estimates with 95% confidence intervals calculated based on standard errors clustered at the course level. Table A.2 shows the corresponding number of observations.

### Results student performance

Table 8 presents descriptive statistics by term for matriculates. Overall, we observe 17,040 matriculates in this period. Their performance does not substantially differ in the summer compared to the winter term. The same holds for the fraction of women per course or the ISEE indicators. Matriculates took on average 2.5 exams in both terms. We observe slightly less bachelor students in the summer term – potentially due to students dropping out over an academic year (given that most degree programs start in the winter). In contrast, we observe more new master students in the summer than in the winter. A reason may be that students finish their bachelor degree during the winter term and proceed with their master studies in the upcoming summer term.

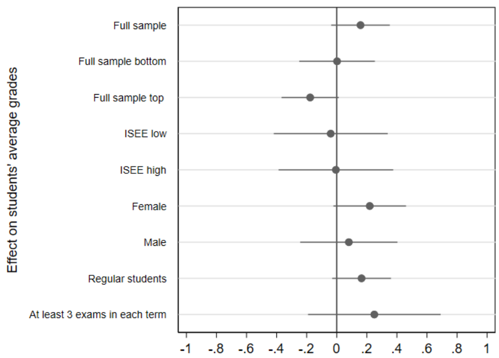
Table 8: Descriptive statistics student performance matriculates, selected controls

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Summer term | | Winter term | | |
| Variable | Mean | Std.Dev. | Mean | Std.Dev. | Difference |
| Average grade | 26.16 | 3.294 | 26.06 | 3.186 | 0.090 |
| Age (years) | 21.70 | 2.224 | 22.52 | 2.093 | -0.820*** |
| Female (dummy) | 0.584 | 0.493 | 0.588 | 0.492 | -0.004 |
| Regular student (dummy) | 1.000 | 0.0149 | 0.991 | 0.0920 | 0.009*** |
| ISEE bottom | 0.265 | 0.441 | 0.259 | 0.438 | 0.006 |
| ISEE medium bottom | 0.254 | 0.435 | 0.261 | 0.439 | -0.007 |
| ISEE medium top | 0.218 | 0.413 | 0.242 | 0.429 | -0.024*** |
| ISEE top | 0.263 | 0.440 | 0.237 | 0.426 | -0.030*** |
| Number of exams per term | 2.478 | 1.319 | 2.508 | 1.192 | -0.030 |
| Bachelor (dummy) | 0.525 | 0.499 | 0.551 | 0.497 | 0.03*** |
| Master (dummy) | 0.336 | 0.472 | 0.299 | 0.458 | 0.037*** |
| Observations | 9,076 | | 7,964 | | 17,040 |

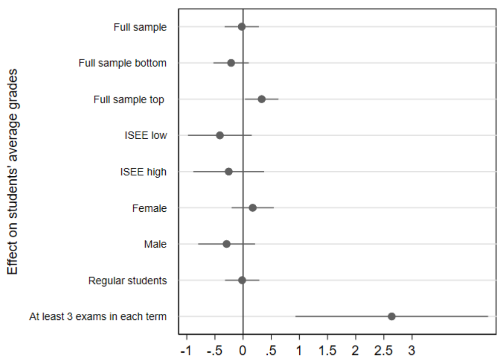
*Notes:* Descriptive statistics refer to academic years 2018/2019 and 2019/2020. Reported differences are based on a regression in the winter term of the selected variables on a period dummy. *, and denote significance at the 10%-, 5%- and 1%-level, respectively. Standard errors clustered at the individual level are used.

We present in Figure 14 the estimation outcome of the specification with control variables for matriculates. As in case of the main analysis, we find no substantial effect of the COVID-19 pandemic on students' average grades. The results suggest no adverse effects for top and bottom students or for students from poorer and richer families. Similarly, we find no gender differences in effects. These insights, thus, support our finding from the main analysis (and from the robustness test for courses not changed) that the transition to online teaching due to the COVID-19 pandemic did not affect student performance substantially. This finding holds for all matriculates as well as for first-year bachelor and master students. A major difference compared to the previous results is that the adverse effect for bottom and top students no longer persists. Further, matriculates from a bachelor course taking at least three exams per term are no longer penalized in terms of grades. In contrast, they experience a non-negligible and statistically significantly positive effect of more than 2.5 grade points. However, this is the only subgroup for that we find a marked effect.

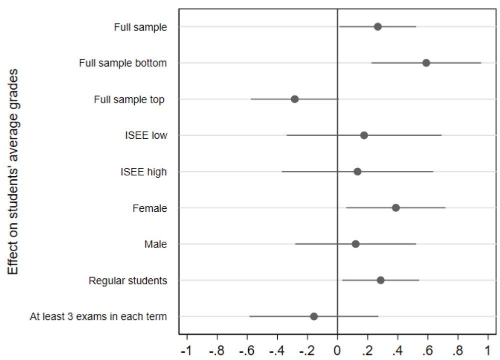
Figure 14: Effects on matriculates' average grades of online teaching due to the COVID-19 pandemic (controls added)




(a) Full sample




(b) Bachelor




(c) Master

*Notes:* Estimates of effects on average grade. Full sample refers to matriculates. The figure shows estimates of the effects on students' average grade from a difference-in-differences specification with controls. Control variables used are gender, age, dummies for being a regular student, ISEE, master, 6-year degree or 5-year degree and year of matriculation. The dependent variable is the student-level average of corrected grades $\tilde{v}_{ist}$. Standard errors clustered at the individual level are used. Point estimates with 95% confidence intervals. Table A.3 shows the corresponding number of observations.

## 6.3 Mandatory courses only

If all courses must be attended by all students, then the distribution of students' reporting styles is the same in all courses and reporting heterogeneity does not bias the relative evaluation of a course. In order to rule out or to reduce reporting bias, we repeat the analysis in this subsection considering only mandatory courses. The experimental setting and estimation approach do not change.

**Results SET**

Table 9 shows descriptive statistics for SET in case of mandatory courses. In total, we have 4,627 obligatory courses in the academic years 2018/2019 and 2019/2020. The indicators as well as the fraction of women per course do neither statistically nor economically significantly vary between the summer and winter term, respectively. In the summer about six questionnaires less per course are filled.

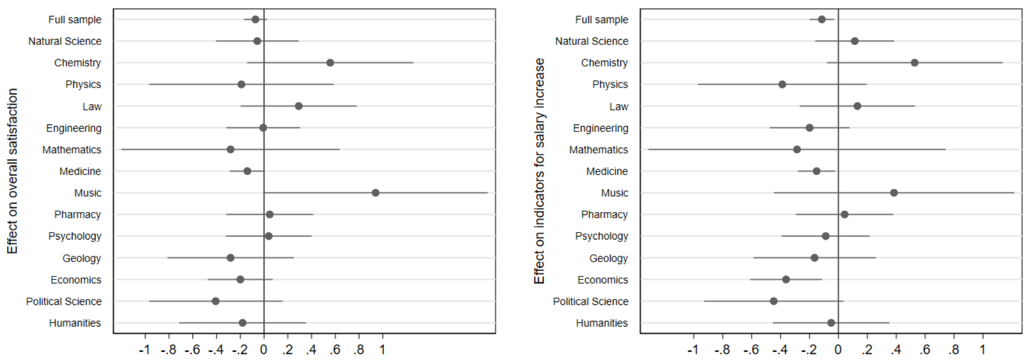
Table 9: Descriptive statistics SET by semester mandatory courses, selected controls

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Summer term | | Winter term | | |
| Variable | Mean | Std.Dev. | Mean | Std.Dev. | Difference |
| Indicator for salary increase | 8.366 | 0.971 | 8.402 | 0.881 | -0.036 |
| Overall satisfaction | 8.098 | 1.113 | 8.151 | 0.995 | -0.053 |
| Number of questionnaires compiled | 37.89 | 45.72 | 44.25 | 52.63 | -6.360*** |
| Women (fraction) | 0.631 | 0.246 | 0.623 | 0.246 | 0.008 |
| Observations | 2,135 | | 2,492 | | 4,627 |

*Notes:* Reported differences are based on a regression of the selected variables on a summer-term dummy. *, and denote significance at the 10%-, 5%- and 1%-level, respectively. Standard errors clustered at the course level are used.

Figure 15 shows the effect of the COVID-19 pandemic on SET in case of mandatory courses only. We find negative and tiny effects for the full sample and both indicators. Moreover, in case of the indicator for overall satisfaction (panel (a)), the point estimates are always statistically insignificant. Indeed, effect heterogeneity among departments is again not very pronounced. To be precise, except for the medicine and economics department in case of the salary-increase-related indicator, we find no statistically significant effect. Thus, only in two out of 14 departments, we find a statistically significant effect. Moreover, the latter is economically small amounting to 0.18 (medicine) or 0.42 (economics) points.

Figure 15: Effect on course SET of online teaching due to the COVID-19 pandemic, mandatory courses (controls added)




(a) Indicator for overall satisfaction (b) Indicator for salary increase

*Notes:* Estimates of effects on SET for mandatory courses and by departments. Full sample refers to the sample of mandatory courses. The figure shows estimates of the effects on SET from a difference-in-differences specification that compares values on SET for the courses taught in the winter and in the summer term. Control variables used are number of questionnaires compiled, female share, course year and lecturer fixed effects. Point estimates with 95% confidence intervals calculated based on standard errors clustered at the course level. Table A.2 shows the corresponding number of observations.

### Results student performance

Table 10 shows descriptive statistics for mandatory courses by term. Considering only mandatory courses leaves us with 41,893 observations. On average, grades are 0.4 points higher in the summer compared to the winter term . We observe similar portions of women and regular students in both terms. The ISEE indicators do also not change substantially over the terms. Further, students take on average 2.8 exams in the summer and 2.7 exams in the winter term. Bachelor students attend slightly more often mandatory courses in the summer (58%) compared to the winter (54%). For master students the opposite holds: about 12% of all students are master students in the summer, while 14% of all students are master students in the winter term.

Figure 16 shows the conditional effects of the COVID-19 pandemic on student

Table 10: Descriptive statistics student performance mandatory courses, selected controls

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Summer term | | Winter term | | |
| Variable | Mean | Std.Dev. | Mean | Std.Dev. | Difference |
| Average grade | 25.86 | 3.226 | 25.44 | 3.405 | 0.420*** |
| Age (years) | 22.81 | 2.210 | 22.91 | 2.214 | -0.10*** |
| Female (dummy) | 0.583 | 0.493 | 0.586 | 0.493 | -0.003 |
| Regular student (dummy) | 0.871 | 0.336 | 0.867 | 0.339 | 0.004* |
| ISEE bottom | 0.253 | 0.435 | 0.242 | 0.429 | 0.011*** |
| ISEE medium bottom | 0.246 | 0.431 | 0.244 | 0.429 | 0.002 |
| ISEE medium top | 0.236 | 0.425 | 0.238 | 0.426 | -0.002 |
| ISEE top | 0.265 | 0.441 | 0.276 | 0.447 | -0.011*** |
| Number of exams per term | 2.771 | 1.424 | 2.670 | 1.408 | 0.101*** |
| Bachelor (dummy) | 0.583 | 0.493 | 0.544 | 0.498 | 0.039*** |
| Master (dummy) | 0.123 | 0.329 | 0.135 | 0.342 | -0.012*** |
| Observations | 21,547 | | 20,346 | | 41,893 |

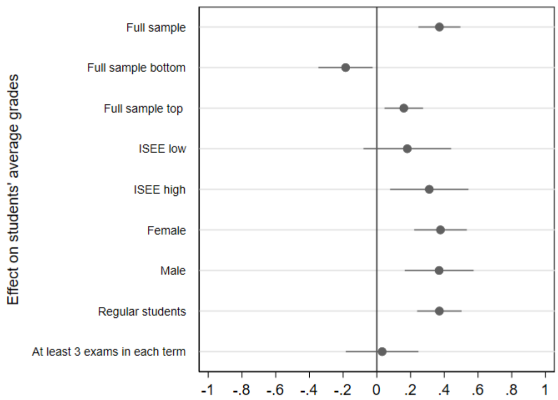
*Notes:* Reported differences are based on a regression of the selected variables on a summer-term dummy. *, and denote significance at the 10%-, 5%- and 1%-level, respectively. Standard errors clustered at the individual level are used.

performance in case of mandatory courses. Overall, the effects are slightly positive and statistically significantly. Yet, a point estimate of 0.32 presents only an effect of 1.2% (given average grades of 25.7 for mandatory courses). The results suggest statistically significant and negative (positive) effects for bottom (top) students. As the confidence bands do not overlap, we find adverse effects along this dimension. However, the effects are quantitatively small. Students from poorer and richer families are slightly positive, though not statistically significantly different affected by the COVID-19 pandemic. Similarly, we find positive effects for both men and women, but no gender differences in effects. Estimated effects for regular students are slightly positive, while we find insignificant effects for students taking at least three exams per term. Overall, the findings do not substantially differ for bachelor and master students or from results from the main analysis. Consequently, this robustness test supports our main insight that the pandemic did not or only marginally affect student performance.

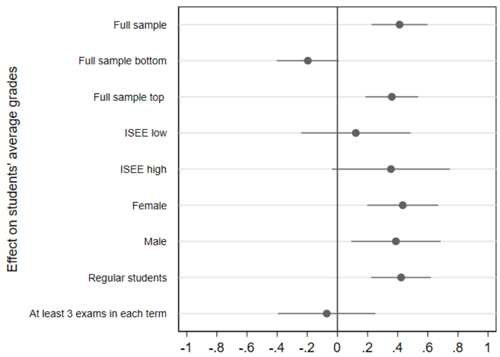
## 6.4 Multple pre-treatment periods

In this section, we use the academic years 2015/2016, 2016/2017, 2017/2018 and 2018/2019 as control group. The experimental design consists again of treatment in

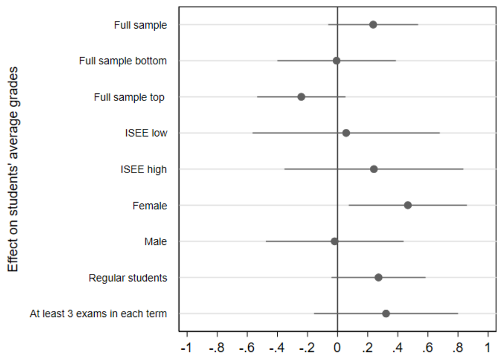
Figure 16: Effects on students' average grades of online teaching due to the COVID-19 pandemic, mandatory courses (controls added)




(a) Full sample




(b) Bachelor




(c) Master

*Notes:* Estimates of effects on average grade. Full sample refers to mandatory courses. The figure shows estimates of the effects on students' average grade from a difference-in-differences specification with controls. Control variables used are gender, age, dummies for being a regular student, ISEE, master, 6-year degree or 5-year degree and year of matriculation. The dependent variable is the student-level average of corrected grades $\tilde{v}_{ist}$. Standard errors clustered at the individual level are used. Point estimates with 95% confidence intervals. Table A.3 shows the corresponding number of observations.

the summer term 2020.

When several pre-treatment periods are available, identification of the treatment effect in a difference-in-differences framework requires an assumption relating dynamics for control and treated group in absence of treatment (Mora and Reggio, 2015). In case of SET, we estimate the following model with flexible common dynamics (see e.g. Mora and Reggio, 2015):

$$
\begin{aligned}
y_{jst}^{k,mc} = \alpha^{k,mc} + \gamma^{k,mc} d_s + \delta^{k,mc} post_t + \beta^{k,mc} d_s * post_t + z_{jst} \lambda^{k,mc} \\
+ \sum_{l=1}^{L} \phi_l^{k,mc} * h_l + \sum_{\tau=2015/2016}^{2019/2020} \zeta_\tau^{k,mc} year_{\tau,t} + w_{jst}^{k,mc}
\end{aligned}
\tag{4}
$$

where $y_{jst}$ is the average overall ($k = 1$) or salary-relevant indicator ($k = 2$) obtained by all students in a specific course $j$ in term $s$ and year $t$. $d$ and $post$ are dummy variables for the summer term and academic year 2019/2020, respectively. The interaction term $d * post$ defines treatment, i.e. summer term 2020. $z$ represents course-level control variables, $h$ are again lecturer dummies and $year$ are academic-year fixed effects, with $year = \mathbb{1}(t = \tau)$ and $post = \mathbb{1}(t \geq 2019/2020)$. $w$ is the corresponding error term. Further, *mc* identifies the coefficient estimates from the analysis with multiple pre-treatment periods.

We estimate the following equation for student performance:

$$
\tilde{v}_{ist}^{mc} = \alpha^{mc} + \gamma^{mc} d_s + \delta^{mc} post_t + \beta^{mc} d_s * post_t + x_{ist} \eta^{mc} + \sum_{\tau=2015/2016}^{2019/2020} \zeta_\tau^{mc} year_{\tau,t} + e_{ist}^{mc}
\tag{5}
$$

where $\tilde{v}_{ist}$ is the average grade obtained by student $i$ in the exams of term $s$ and academic year $t$ depurated from lecturer fixed effects (see section 3 for details), $x$ represents student-level control variables, $year$ are academic-year fixed effects and $e$ is the corresponding error term. As in case of the SET in equation (4), $d$ represents a dummy variable for the summer term and $post$ is a dummy for academic year 2019/2020. The interaction term $d * post$ defines again the treatment.

**Results SET**

Table 11 shows that descriptive statistics for SET indicators and related control variables by term for multiple pre-treatment periods. All indicators except number of questionnaires compiled do not substantially differ by term. Compared to the main analysis, students evaluate more questionnaires per course in the summer term in this sample. For comparison, in the main analysis, students compiled on average 36 questionnaires in the summer and 21 in the winter.

Table 11: Descriptive statistics SET by term multiple pre-treatment periods, selected controls

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Summer term | | Winter term | | |
| Variable | Mean | Std.Dev. | Mean | Std.Dev. | Difference |
| Indicator for salary increase | 8.515 | 1.004 | 8.494 | 1.112 | 0.021 |
| Overall satisfaction | 8.249 | 1.138 | 8.225 | 1.259 | 0.024 |
| Number questionnaires compiled | 30.680 | 45.350 | 20.960 | 32.850 | 9.720*** |
| Women (fraction) | 0.594 | 0.277 | 0.581 | 0.301 | 0.013 |
| Observations | 12,705 | | 9,982 | | 22,687 |

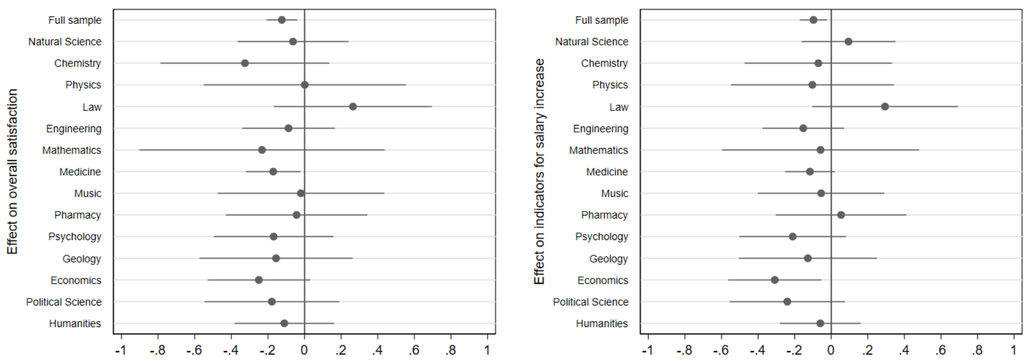
*Notes:* Reported differences are based on a regression of the selected variables on a summer-term dummy. *, and denote significance at the 10%-, 5%- and 1%-level, respectively.

For the analysis of the effect of the COVID-19 pandemic on SET, we do not reject the null hypothesis of common pre-dynamics in case of both the overall-satisfaction indicator (5.86 p-value: 0.12) and the salary-related indicator (7.32 p-value: 0.06).

Figure 17 shows the results for SET conditional on general observable characteristics. As in case of the main analysis, we find no marked effect of the COVID-19 pandemic on overall satisfaction or the indicator for salary increase. This finding holds for the full sample as well as the distinct departments. Now only SET in the medicine and economics department are statistically significantly negatively affected. That is, in only two out of 14 departments, we find statistically significant effects. In case of the overall satisfaction indicator, SET in the department of medicine are 0.2 points lower. The latter represents a reduction of 2.4% (given an average indicator of 8.2). In the economics department, the salary-related indicator is evaluated 0.3 points lower due the pandemic. That is, the indicator is negatively affected by 3.5% (given an average indicator of 8.5). This effect is, thus, again relatively small.

To sum up, also in case of multiple pre-treatment periods, we observe no pronounced heterogeneity in effects across departments. In fact, the analysis suggests statistically significant effects in only two out of 14 departments. Overall, this robustness test confirms again the results from the main analysis of modest to no effects.

Figure 17: Effect on course SET of online teaching due to the COVID-19 pandemic, multiple pre-treatment periods (controls added)




(a) Indicator for overall satisfaction (b) Indicator for salary increase

*Notes:* Estimates of effects on SET for the multiple pre-treatment periods and by departments. Full sample refers to academic years 2015/2016 - 2018/2019 (control group) and 2019/2020 (treatment group). The figure shows estimates of the effects on SET from a difference-in-differences specification that compares values on SET for the courses taught in the winter and in the summer term. Control variables used are number of questionnaires compiled, female share, course year and lecturer fixed effects. Point estimates with 95% confidence intervals calculated. Table A.2 shows the corresponding number of observations.

### Results student performance

Table 12 shows descriptive statistics for student performance and several time periods. In total, we have now 145,966 observations. On average students perform about 0.4 grade points better in the summer compared to the winter term. The latter is comparable to descriptive statistics from previous analyses. The differences are generally small such that balancing is again not a problem. Note that the differences are statistically significant throughout given the large number of observations (145,966).

For the analysis on student performance, we do not reject the null hypothesis of

Table 12: Descriptive statistics student performance multiple pre-treatment periods, selected controls

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Summer term | | Winter term | | |
| Variable | Mean | Std.Dev. | Mean | Std.Dev. | Difference |
| Average grade | 26.06 | 3.192 | 25.62 | 3.335 | 0.440*** |
| Age (years) | 22.88 | 2.232 | 22.92 | 2.269 | -0.040*** |
| Female (dummy) | 0.575 | 0.494 | 0.578 | 0.494 | -0.003*** |
| Regular student (dummy) | 0.725 | 0.447 | 0.713 | 0.452 | 0.012*** |
| ISEE bottom | 0.255 | 0.436 | 0.253 | 0.434 | 0.002*** |
| ISEE medium bottom | 0.251 | 0.433 | 0.249 | 0.432 | 0.002*** |
| ISEE medium top | 0.373 | 0.484 | 0.383 | 0.486 | -0.010*** |
| ISEE top | 0.121 | 0.3261 | 0.118 | 0.323 | 0.003*** |
| Number of exams per term | 2.782 | 1.422 | 2.773 | 1.511 | 0.009*** |
| Bachelor (dummy) | 0.555 | 0.497 | 0.548 | 0.498 | 0.007*** |
| Master (dummy) | 0.186 | 0.389 | 0.193 | 0.394 | -0.007*** |
| Observations | 73,289 | | 72,677 | | 145,966 |

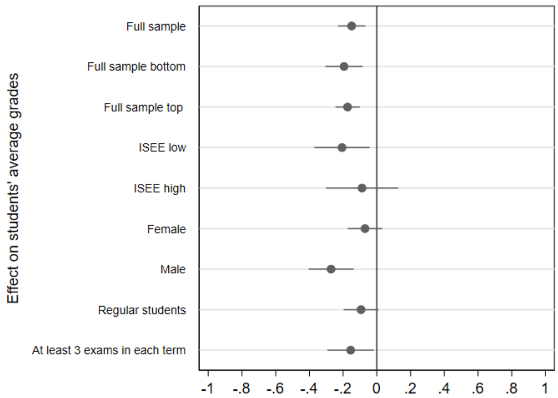
*Notes:* Reported differences are based on a regression of the selected variables on a summer-term dummy. *, and denote significance at the 10%-, 5%- and 1%-level, respectively.

a Wald test of joint significance of all interactions of pre-treatment time dummies and the treatment dummy at a 10% significance level: 5.17, p-value: 0.16. Figure 18 shows the estimation outcome in case of multiple pre-treatment periods. We find generally negative, small, though statistically significant effects. The results suggest no adverse effects for bottom and top students, for students from poorer and richer families or for men and women (overlapping confidence intervals). The estimated effects are driven by bachelor students. Qualitatively, the effects are small throughout amounting at most to 2.4% in case of male bachelor students (point estimate: 0.6, average grade: 25). That is, we find again no marked effect of the COVID-19 pandemic on student performance.

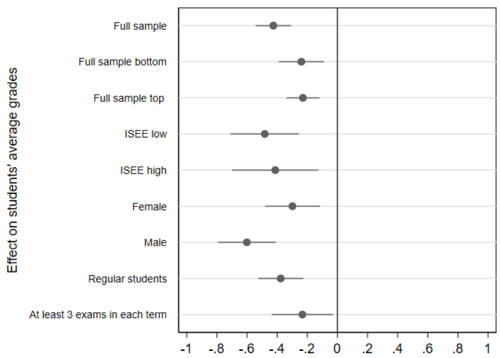
# 7 Discussion

We find only negligible effects of the transition to online teaching due to the first wave of the COVID-19 pandemic at the University of Pavia. This finding holds for both educational quality (SET) and student performance (students' average grades). Overall, SET are slightly negatively affected, while students' average grades are slightly positively affected. Given these surprising results, below we present a detailed discussion of the factors that, in contrast, may have been expected to affect

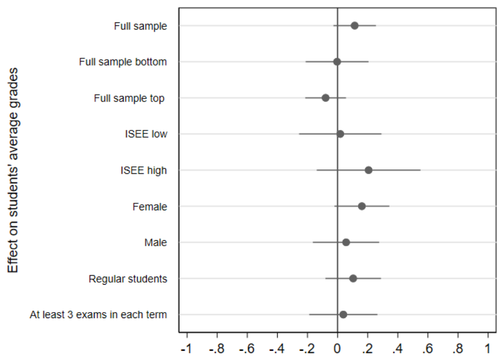
Figure 18: Effects on students' average grades of online teaching due to the COVID-19 pandemic multiple pre-treatment periods (controls added)




(a) Full sample




(b) Bachelor




(c) Master

*Notes:* Estimates of effects on average grade. Full sample refers to academic year 2015/2016 - 2018/2019 (control group) and 2019/2020 (treatment group). The figure shows estimates of the effects on students' average grade from a difference-in-differences specification with controls. Control variables used are gender, age, dummies for being a regular student, ISEE, master, 6-year degree or 5-year degree and year of matriculation. The dependent variable is the student-level average of corrected grades $\tilde{v}_{ist}$. Standard errors clustered at the individual level are used. Point estimates with 95% confidence intervals. Table A.3 shows the corresponding number of observations.

the outcomes of students and lecturers over the period in question and of how we dealt with them in our empirical analysis.

We find no significant variation in the performance of both students and lecturers due to the COVID-19 pandemic. Yet, the occurrence (or the expectation) of a deterioration of health and economic conditions for themselves, family members and relates may have been expected to alter the behaviour of both students and lecturers. For students, these effects, if any, should exhibit some relation with their family background and/or the individual ability. To test for these natural conjectures, first, we conduct the analysis for top- and bottom-performing students and, second, for students from low- and high-income families. Our results suggest that the COVID-19 pandemic does not widen existing gaps in student performance along these dimensions.[7]

Apart from affecting average grades of students, the pandemic may have led to more (bottom) students falling behind by passing less exams. As we have no student-level data along this dimension, we look descriptively at potential effects of the pandemic on exam failure rates. However, we find a slight (about 2 percentage points) decline in the rate of exams failed in the summer 2020 compared to previous years. That is, the pandemic did not lead to more students falling behind.

Moreover, students may have taken less exams in the summer 2020, self-selecting themselves into subjects with higher probability of success (maybe because they were particularly interested in them). However, the data (e.g. Table 4) suggests no difference in the average number of exams taken during the treatment and the results do not change when using an indicator for overall satisfaction free of interest in subject (as shown in Figure B.2 in Appendix B). In addition, if we consider only mandatory courses that reduce potential self-selection bias in subjects. We find that the pandemic did not affect student performance or the assessment of educational quality.

Since we have neither data on family income nor on health conditions on the lecturer side, we run our regressions with and without lecturer fixed effects finding basically no difference. Further, lecturers may have lowered the level of difficulty of exams or graded student outcomes more generously in order to compensate them for the special situation. Note that at the University of Pavia grades are not normalized but lecturers stick to the same grading scale. That is, lecturers do not systematically

[7]These findings are in contrast to findings from the literature of the COVID-19 pandemic on schools (among others Agostinelli et al., 2020; Aucejo et al., 2020; Bacher-Hicks et al., 2021; Chetty et al., 2020).

attribute to the best-performing student in the course the highest grade. To test for lecturer heterogeneity in exams and in allocation of grades, we use – again – lecturer fixed effects. As the results do substantially not change with or without lecturer fixed effects, we do not find any evidence supporting this concern.

To account for the fact that students may have evaluated the courses more positively during the COVID-19 pandemic (for instance, to reward the effort made by lecturers to react to the new teaching organization), we run the analysis considering only matriculates who had no past relationship with the lecturers and no comparison with a previous year. Again, we find no effect supporting a positive evaluation bias of students.

Much attention has been devoted to whether female lecturers receive better or worse evaluation than their male counterparts (Boring, 2017; Wagner et al., 2016). We account for this issue by including lecturer fixed effects in the regression and by considering separately male and female students (e.g. Engzell et al., 2020, found different effects for girls and boys). The results suggest that the transition to online teaching did not lead to significant gender differences in higher education.

# 8 Conclusion

This paper suggests that the first wave of the COVID-19 pandemic did not represent a threat to higher education. We estimate the causal effect of the transition to online teaching due to the first wave of the COVID-19 pandemic on educational quality (SET) and student performance (average grades) at the University of Pavia (Lombardy, Italy). Our study is – to our best knowledge – the first that considers the effects of the pandemic on universities using administrative university data. The rich data set allows us to control for various socioeconomic dimensions such as family income or gender.

The transition to online teaching in Lombardy and the North of Italy in spring 2020 coincided with the beginning of the summer term. The latter permits us to identify the causal effect of the COVID-19 pandemic on higher education. The estimation approach consists of a standard difference-in-differences setup, with the transition to online teaching due to the pandemic being the treatment. We run the analysis separately for SET indicators and average student grades. Further, we look at associations of COVID-19 on graduation grades and exam failure rates.

The results suggest that the first wave of the COVID-19 pandemic did not represent a threat to higher education in terms of educational quality and student performance. In fact, we find no substantial effect of the COVID-19 pandemic neither on SET nor on students' grades. The general tendency of the estimated effects on educational quality was slightly negative, while that on student performance was slightly positive. Our results are robust across different departments, subsamples (courses not changed, matriculates, mandatory courses) and along various dimensions (gender, rich/poor family background, top/bottom students). Our findings differ from those of the emerging COVID-19 literature on schools that identified significant negative affects along various dimensions (Agostinelli et al., 2020; Bacher-Hicks et al., 2021; Engzell et al., 2020). Similarly, our results differ from those of Aucejo et al. (2020) that found pronounced negative effects of the pandemic for university students' outcome and expectations based on survey data. In contrast to the literature of the COVID-19 pandemic on education (Agostinelli et al., 2020; Aucejo et al., 2020; Bacher-Hicks et al., 2021; Chetty et al., 2020), we also find no adverse effects for top or bottom students, for students from poorer or richer families or for men and women.

Moreover, we consider multiple pre-treatment periods in order to make statements about common trends and to exclude that the results are driven by a specific control group. Our main findings do again not change. Descriptive evidence on exam failure rates over time suggests that not more students failed exams due to the pandemic. Similarly, graduation was not affected by COVID-19. A caveat of this study is that we have only descriptive evidence on the effects of the pandemic on exam failures and graduation. Further, we can only make statements about the short run. That is, this paper is silent about the effect of the COVID-19 pandemic in the medium and long run.

Overall, our analysis suggests that there were no or only modest effects of the COVID-19 pandemic on SET and average grades. Thus, in the short run, higher education was not affected by the pandemic. As a policy implication, these findings suggest that concerns of increasing educational inequality do not apply for universities. An interesting open research topic is to verify the results also for other universities in Italy and worldwide.

# References

Agostinelli, F., Doepke, M., Sorrenti, G., Zilibotti, F., 2020. When the Great Equalizer Shuts Down: Schools, Peers, and Parents in Pandemic Times. IZA Discussion Paper No. 13965.

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

Bacher-Hicks, A., Goodman, J., Mulhern, C., 2021. Inequality in household adaptation to schooling shocks: Covid-induced online learning engagement in real time. Journal of Public Economics 193, 104345.

Bahasoan, A.N., Ayuandiani, W., Mukhram, M., Rahmat, A., 2020. Effectiveness of online learning in pandemic covid-19. International Journal of Science, Technology and Management 1, 100–106.

Bargain, O., Etienne, A., Melly, B., 2018. Public sector wage gaps over the long-run: Evidence from panel administrative data. IZA Discussion Paper No. 11924.

Bertoni, M., Rettore, E., Rocco, L., 2020. If (My) 6 Was (Your) 9: Reporting Heterogeneity in Student Evaluations of Teaching. IZA Discussion Paper No. 13565.

Bonaccolto-Töpfer, M., Castagnetti, C., Prümer, S., 2021. Does it pay to go public? Understanding the public-private sector wage gap in Germany. FAU Working Paper 116.

Boring, A., 2017. Gender biases in student evaluations of teachers. Journal of Public Economics 145, 27–41.

Canay, I.A., 2011. A simple approach to quantile regression for panel data. The Econometrics Journal 14, 368–386.

Castagnetti, C., Giorgetti, M.L., 2019. Understanding the gender wage-gap differential between the public and private sectors in Italy: A quantile approach. Economic Modelling 78, 240–261.

Chetty, R., Friedman, J., Hendren, N., Stepner, M., et al., 2020. How did COVID-19 and Stabilization Policies affect Spending and Employment ? A New Real-Time Economic Tracker Based on Private Sector Data. NBER Working Paper 27431.

Engzell, P., Frey, A., Verhagen, M., 2020. Pre-analysis plan for: Learning inequality during the COVID-19 pandemic. SocArXiv. Center for Open Science.

Kahn, L.B., 2010. The long-term labor market consequences of graduating from college in a bad economy. Labour economics 17, 303–316.

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

Lalla, M., Facchinetti, G., Mastroleo, G., 2005. Ordinal scales and fuzzy set systems to measure agreement: An application to the evaluation of teaching activity. Quality and Quantity 38, 577–601.

Mishra, L., Gupta, T., Shree, A., 2020. Online teaching-learning in higher education during lockdown period of COVID-19 pandemic. International Journal of Educational Research Open 1, 100012.

Mora, R., Reggio, I., 2015. didq: A command for treatment-effect estimation under alternative assumptions. The Stata Journal 15, 796–808.

Oreopoulos, P., Von Wachter, T., Heisz, A., 2012. The short-and long-term career effects of graduating in a recession. American Economic Journal: Applied Economics 4, 1–29.

Susskind, D., Vines, D., 2020. The economics of the covid-19 pandemic: an assessment. Oxford Review of Economic Policy 36, 1–13.

Valmon, 2020. Valutazione e monitoraggio. `http://www.valmonsrl.it/`. Accessed 29-03-2021.

Wagner, W., Göllner, R., Werth, S., Voss, T., Schmitz, B., Trautwein, U., 2016. Student and teacher ratings of instructional quality: Consistency of ratings over time, agreement, and predictive power. Journal of Educational Psychology 108, 705–721.

# A Further descriptives and number of observations

Table A.1 shows the relation between overall satisfaction and the other indicators. The indicators most related with overall satisfaction are those related to lecturer's motivation, teaching and effectiveness. In contrast, indicators for organizational matters are less relevant. Overall, these items explain a substantial share of the variation in overall satisfaction as the R-squared is 0.8. These associations are in line with e.g. Bertoni et al. (2020). Consequently, we – as well as many university administrations – consider overall satisfaction as a reasonable indicator to analyze teaching quality (Bertoni et al., 2020).

Table A.1: Overall satisfaction and its covariates, OLS estimates

| | (1) Overall satisfaction |
|---|---|
| Lecturer's ability to motivate the class | 0.333*** |
| | (0.017) |
| Lecturer teaches in clear way | 0.248*** |
| | (0.017) |
| Lecturer is available for clarifications | -0.012 |
| | (0.019) |
| Clear presentation of learning objectives | 0.100*** |
| | (0.023) |
| Quality of lecture notes/reference books | 0.132*** |
| | (0.015) |
| Sufficient prerequisites | 0.034** |
| | (0.014) |
| Clear presentation of the exam rules from the beginning | 0.040*** |
| | (0.012) |
| Lecturer present during office hours | 0.007 |
| | (0.004) |
| Workload is consistent with the ECTS | 0.084*** |
| | (0.011) |
| Your interest in the subject | 0.060*** |
| | (0.013) |
| Timetables respected | -0.262*** |
| | (0.043) |
| Constant | 0.062 |
| | (0.214) |
| Observations | 7,809 |
| R-squared | 0.801 |

Standard errors clustered at the course level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.2: Number of observations – SET

| Sample | (1)<br>Treated (before) | (2)<br>Not treated (before) | (3)<br>Total |
|---|---|---|---|
| Full sample | 3,779 (1,838) | 4,030 (1,982) | 7,809 |
| Natural science | 226 (114) | 280 (134) | 506 |
| Chemistry | 94 (48) | 89 (43) | 183 |
| Physics | 98 (48) | 119 (55) | 217 |
| Law | 139 (70) | 123 (56) | 262 |
| Engineering | 462 (231) | 531 (265) | 993 |
| Mathematics | 74 (39) | 69 (33) | 143 |
| Medicine | 1,019 (481) | 1,099 (532) | 2,118 |
| Music | 160 (77) | 164 (88) | 324 |
| Pharmacy | 139 (72) | 158 (76) | 297 |
| Psychology | 290 (134) | 290 (141) | 580 |
| Geology | 136 (68) | 122 (61) | 258 |
| Economics | 252 (118) | 299 (146) | 551 |
| Political science | 236 (124) | 223 (113) | 459 |
| Humanities | 454 (214) | 464 (239) | 918 |
| Courses not changed | 3,080 (1,511) | 3,494 (1,730) | 6,574 |
| Matriculates | 1,688 (828) | 1,609 (788) | 3,297 |
| Obligatory courses | 2,135 (1,087) | 2,492 (1,255) | 4,627 |
| Multiple pre-treatment periods | 12,705 (10,764) | 9,982 (8,143) | 22,687 |

Table A.3: Number of observations – Average grades

| Sample | (1)<br>Treated (before) | (2)<br>Not treated (before) | (3)<br>Total |
|---|---|---|---|
| Full sample | 28,245 (11,771) | 26,363 (11,771) | 54,608 |
| Full sample bottom | 7,457(3,178) | 8,399 (3,403) | 15,856 |
| Full sample top | 8,638 (3,417) | 7,109 (2,804) | 15,747 |
| ISEE low | 7,239 (2,919) | 6,548 (2,557) | 13,787 |
| ISEE high | 7,374 (3,117) | 7,062 (2,803) | 14,436 |
| Female | 16,534 (6,871) | 15,568 (6,176) | 32,102 |
| Male | 11,711 (4,900) | 10,795 (4,259) | 22,506 |
| Regular students | 24,244 (9,831) | 22,504 (8,625) | 46,748 |
| At least three exams each term | 7,743 (3,329) | 6,872 (2,646) | 14,615 |
| Bachelor students | 15,738 (6,533) | 14,318 (5,661) | 30,056 |
| Master students | 5,425 (2,071) | 5,192 (1,845) | 10,617 |
| Natural science | 2,326 (916) | 2,186 (801) | 4,512 |

*Continued on next page*

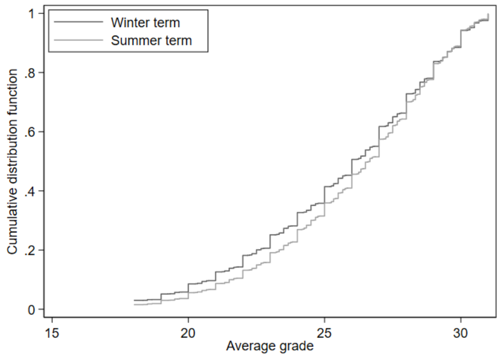
| Sample | (1)<br>Treated (before) | (2)<br>Not treated (before) | (3)<br>Total |
|---|---|---|---|
| Natural science bottom | 516 (225) | 616 (234) | 1,132 |
| Natural science top | 656 (211) | 578 (210) | 1,234 |
| Chemistry | 560 (221) | 514 (195) | 1,074 |
| Chemistry bottom | 131 (55) | 144 (54) | 275 |
| Chemistry top | 152 (47) | 147 (64) | 299 |
| Physics | 366 (134) | 367 (135) | 733 |
| Physics bottom | 93 (37) | 104 (41) | 197 |
| Physics top | 120 (43) | 89 (29) | 209 |
| Law | 1,413 (605) | 1,322 (546) | 2,735 |
| Law bottom | 339 (137) | 417 (166) | 756 |
| Law top | 433 (206) | 313 (133) | 746 |
| Engineering | 4,073 (1,700) | 3,571 (1,421) | 7,644 |
| Engineering bottom | 1,025 (430) | 1,184 (463) | 2,209 |
| Engineering top | 1,436 (608) | 1,002 (439) | 2,438 |
| Mathematics | 271 (106) | 268 (101) | 539 |
| Mathematics bottom | 67 (25) | 91 (30) | 158 |
| Mathematics top | 81 (28) | 68 (21) | 149 |
| Medicine | 5,880 (2,524) | 5,393 (2,061) | 11,273 |
| Medicine bottom | 1,551 (690) | 1,628 (623) | 3,179 |
| Medicine top | 1,810 (681) | 1,466 (589) | 63,276 |
| Music | 365 (150) | 347 (131) | 712 |
| Music bottom | 97 (47) | 90 (39) | 187 |
| Music top | 128 (41) | 106 (36) | 234 |
| Pharmacy | 2,335 (1,026) | 2,340 (1,032) | 4,675 |
| Pharmacy bottom | 618 (297) | 811 (362) | 1,429 |
| Pharmacy top | 787 (302) | 679 (299) | 1,466 |
| Psychology | 1,254 (512) | 1,183 (440) | 2,437 |
| Psychology bottom | 286 (117) | 432 (163) | 718 |
| Psychology top | 479 (170) | 346 (138) | 825 |
| Geology | 699 (292) | 654 (265) | 1,353 |
| Geology bottom | 161 (80) | 239 (107) | 400 |
| Geology top | 259 (99) | 155 (60) | 414 |
| Economics | 3,480 (1,384) | 3,282 (1,279) | 6,762 |
| Economics bottom | 919 (353) | 1,018 (427) | 1,937 |

*Continued on next page*

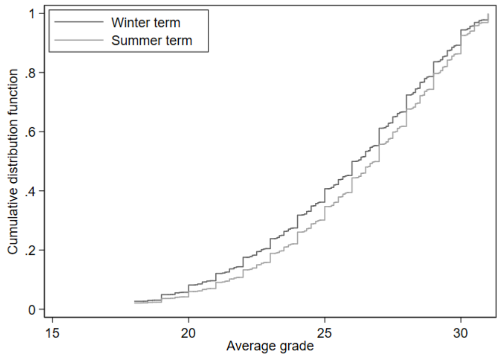
| Sample | (1)<br>Treated (before) | (2)<br>Not treated (before) | (3)<br>Total |
|---|---|---|---|
| Economics top | 953 (400) | 928 (323) | 1,881 |
| Political science | 2,814 (1,191) | 2,660 (1,078) | 5,474 |
| Political science bottom | 694 (305) | 846 (345) | 1,540 |
| Political science top | 849 (349) | 798 (333) | 1,647 |
| Humanities | 2,409 (1,010) | 2,276 (950) | 4,685 |
| Humanities bottom | 630 (286) | 654 (290) | 1,284 |
| Humanities top | 791 (316) | 694 (282) | 1,485 |
| Courses not changed | 25,397 (10,597) | 24,212 (9,697) | 49,609 |
| Matriculates | 9.076 (4,138) | 7,964 (3,660) | 17,040 |
| Obligatory courses | 21,547 (9,423) | 20,346 (8,206) | 41,893 |
| Multiple pre-treatment periods | 73,289 (56,815) | 72,677 (60,848) | 145,966 |

Figure A.1 shows the cumulative distribution function of average grades by semester and academic year. The differences in distribution of average grades between the winter and summer term is not very pronounced. This finding holds for the full sample as well as for bachelor and master students.

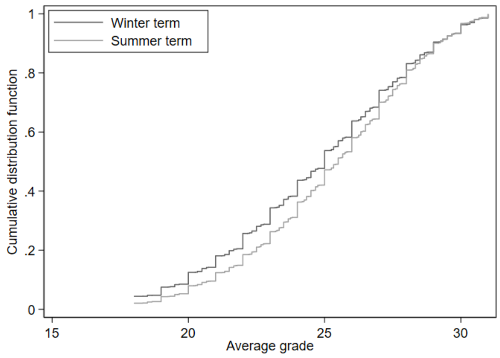
Figure A.1: Cumulative distribution function of average grades by term and academic year



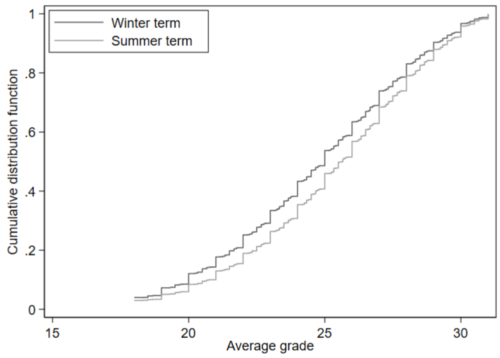

(a) Full sample 2018/2019



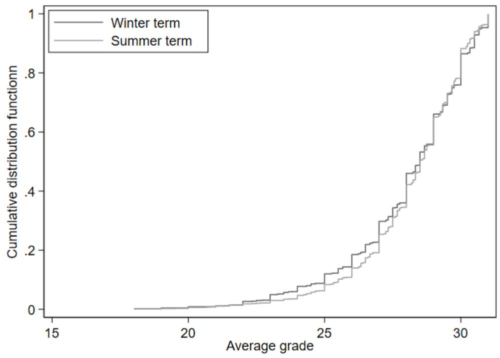

(b) Full sample 2019/2020



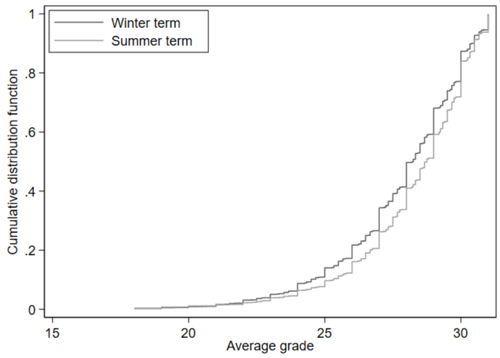

(c) Bachelor 2018/2019




(d) Bachelor 2019/2020




(e) Master 2018/2019

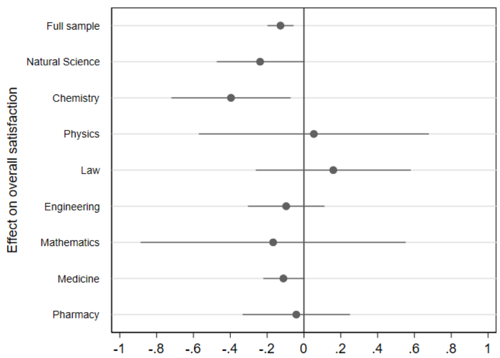



(f) Master 2019/2020

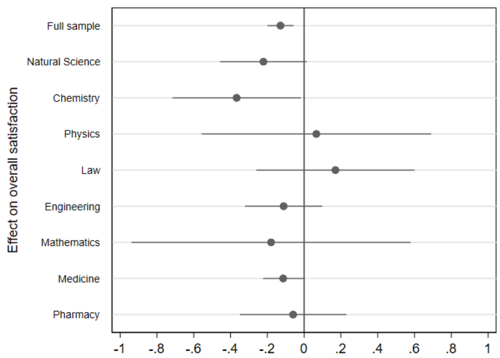
# B SET free of interest in subject

Figure B.2 shows the estimation outcome when using overall satisfaction depurated from interest in subject as dependent variable. Again, we find no significant effect of the COVID-19 pandemic on overall satisfaction and thus on teaching quality. This finding holds in both specifications with and without control variables.

Figure B.2: Effect on course SET of online teaching due to the COVID-19 pandemic, full sample accounting for interest in subject




(a) Without controls

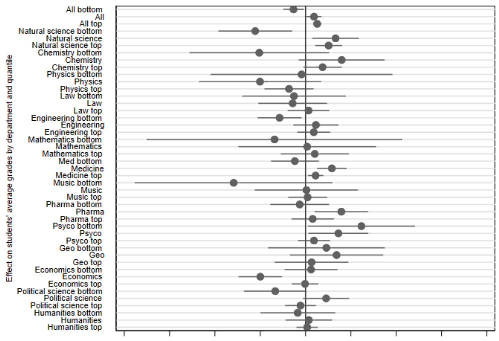



(b) With controls

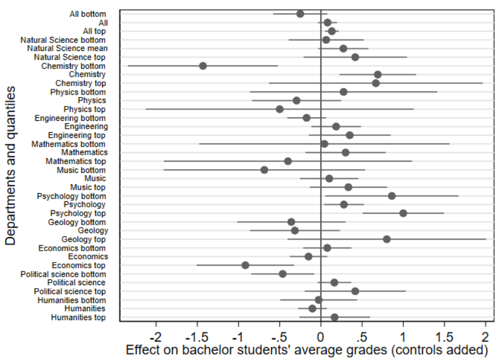
*Notes:* Estimates of effects on SET for the full sample and by departments. Dependent variable is the residual from a regression of overall satisfaction on indicator for interest in subject. The figure shows estimates of the effects on SET from a difference-in-differences specification that compares values on SET for the courses taught in the winter and summer term. Point estimates with 95% confidence intervals calculated based on standard errors clustered at the course level. Table A.2 shows the corresponding number of observations.

# C Results without control variables

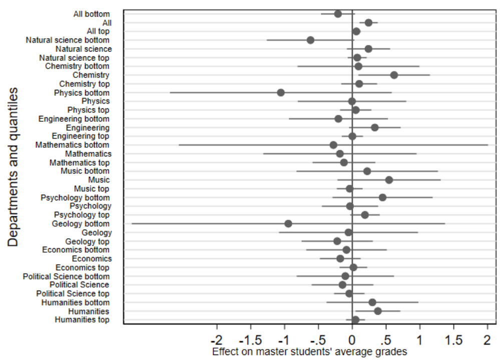
Figure C.3: Effect on average grades of online teaching due to the COVID-19 pandemic, (no controls added)



(a) Full sample



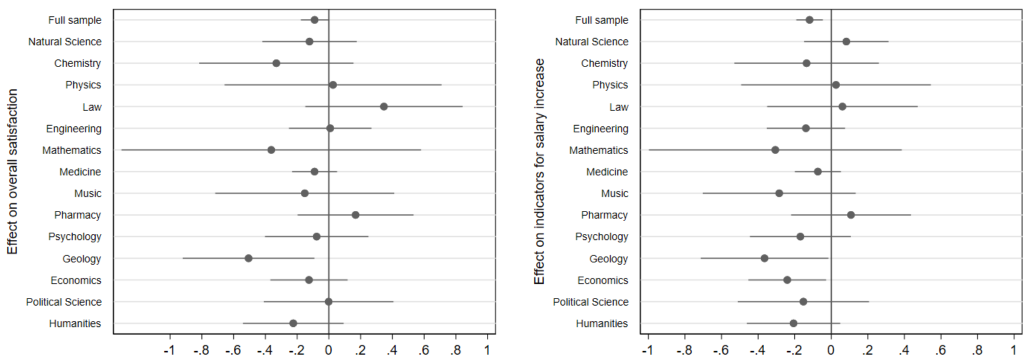

(b) Bachelor




(c) Master

*Notes:* The figure shows estimates of the effects on students' average grade from a difference-in-differences specification without controls. Point estimates with 95% confidence intervals calculated based on standard errors clustered at the individual level. Table A.3 shows the corresponding number of observations.

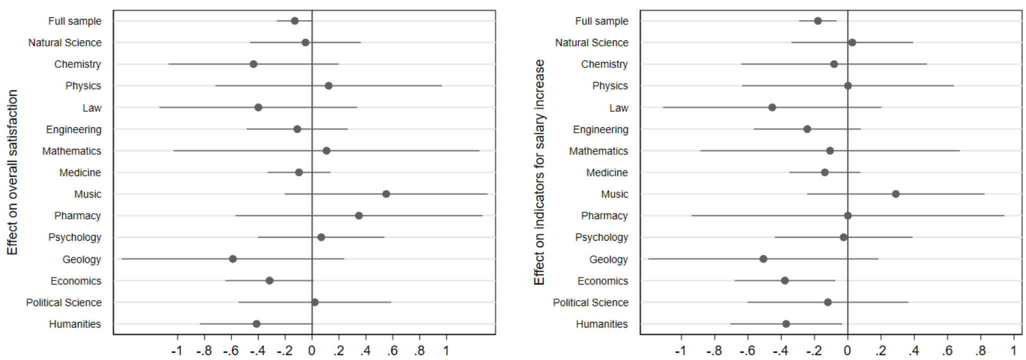
Figure C.4: Effect on course SET of online teaching due to the COVID-19 pandemic, courses not changed (no controls added)




(a) Indicator for overall satisfaction (b) Indicator for salary increase

*Notes:* Estimates of effects on SET for courses not changed and by departments. Full sample refers to courses not changed. The figure shows estimates of the effects on SET from a difference-in-differences specification that compares values on SET for the courses taught in the winter and summer term. Point estimates with 95% confidence intervals calculated based on standard errors clustered at the course level. Table A.2 shows the corresponding number of observations.

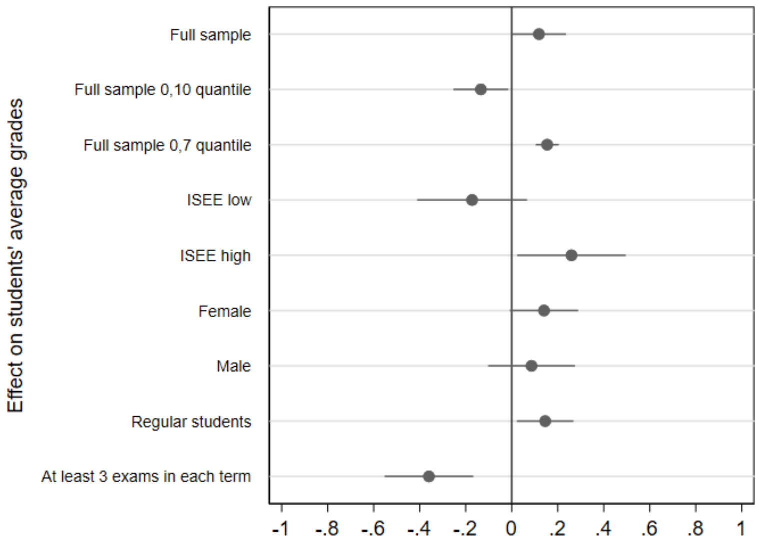
Figure C.6: Effect on course SET of online teaching due to the COVID-19 pandemic, matriculates (no controls added)



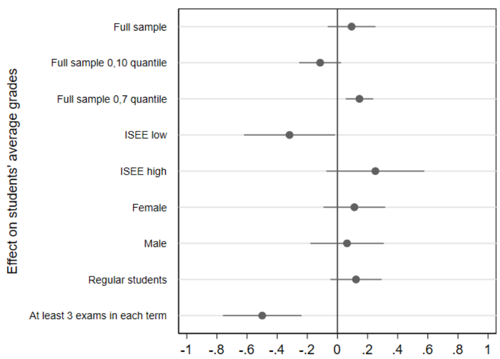

(a) Indicator for overall satisfaction (b) Indicator for salary increase

*Notes:* Estimates of effects on SET for matriculates and by departments. Full sample refers to matriculates. The figure shows estimates of the effects on SET from a difference-in-differences specification that compares values on SET for the courses taught in the winter and summer term. Point estimates with 95% confidence intervals calculated based on standard errors clustered at the course level. Table A.2 shows the corresponding number of observations

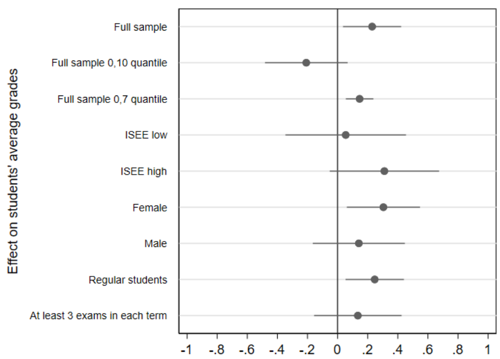
Figure C.5: Effects on average grades of online teaching due to COVID-19 pandemic, courses not changed (no controls added)




(a) Full sample



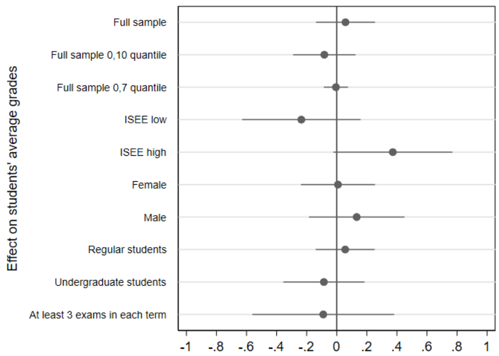

(b) Bachelor



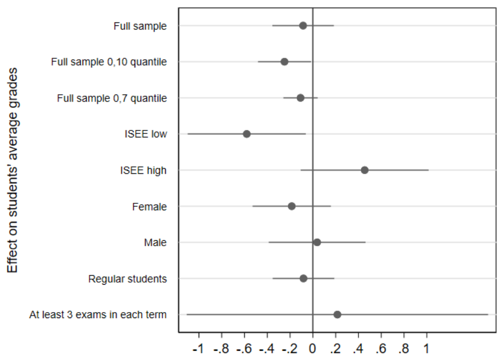

(c) Master

*Notes:* Estimates of effects on average grade. Full sample refers to courses not changed. The figure shows estimates of the effects on students' average grade from a difference-in-differences specification without controls. Point estimates with 95% confidence intervals calculated based on standard errors clustered at the individual level. Table A.3 shows the corresponding number of observations.

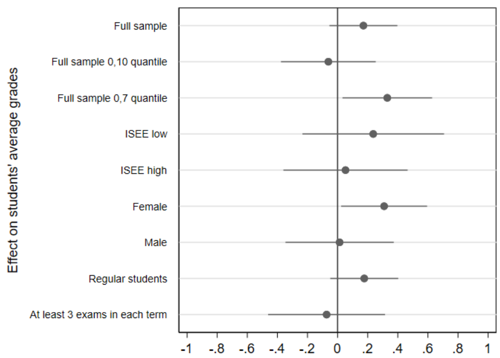
Figure C.7: Effects on matriculates' average grades of online teaching due to the COVID-19 pandemic (no controls added)




(a) Full sample



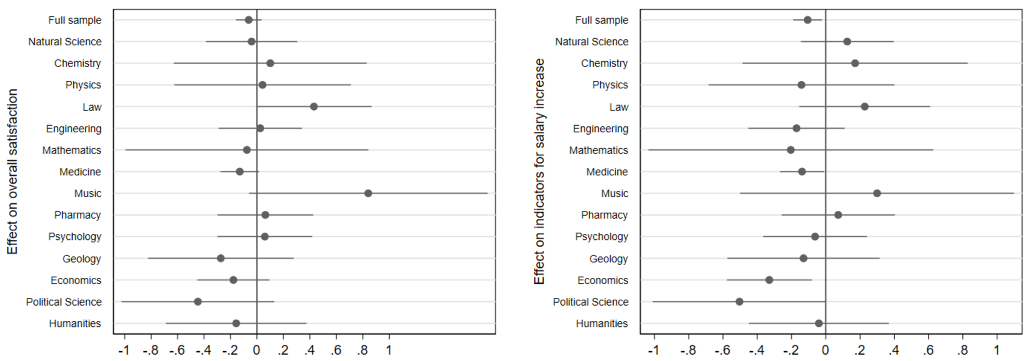

(b) Bachelor




(c) Master

*Notes:* Estimates of effects on average grade. Full sample refers to matriculates. The figure shows estimates of the effects on students' average grade from a difference-in-differences specification without controls. Point estimates with 95% confidence intervals calculated based on standard errors clustered at the individual level. Table A.3 shows the corresponding number of observations.

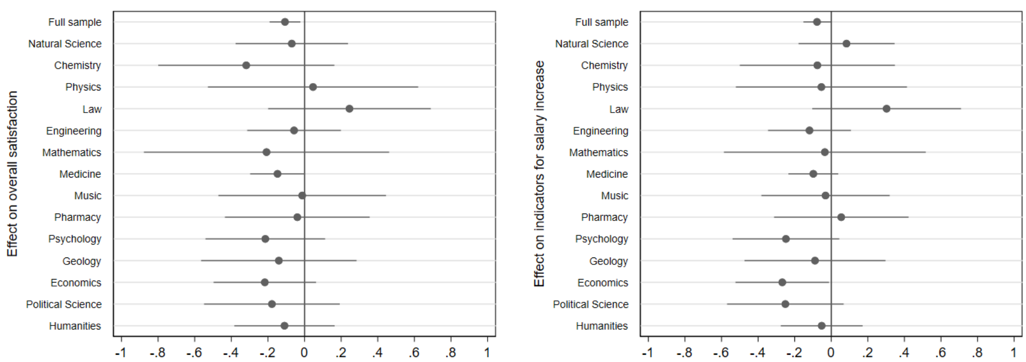
Figure C.8: Effect on course SET of online teaching due to the COVID-19 pandemic, mandatory courses (no controls added)




(a) Indicator for overall satisfaction (b) Indicator for salary increase

*Notes:* Estimates of effects on SET for the mandatory courses and by departments. Full sample refers to mandatory courses. The figure shows estimates of the effects on SET from a difference-in-differences specification that compares values on SET for the courses taught in the winter and summer term. Point estimates with 95% confidence intervals calculated based on standard errors clustered at the course level. Table A.2 shows the corresponding number of observations

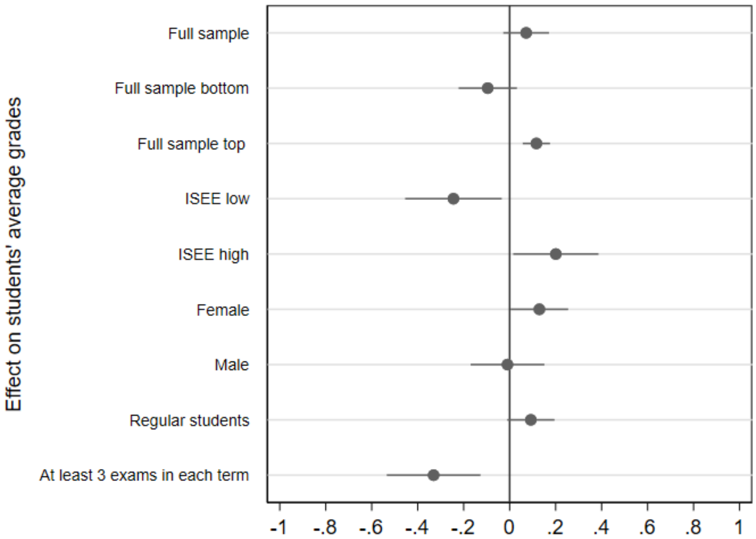
Figure C.10: Effect on course SET of online teaching due to the COVID-19 pandemic, multiple pre-treatment periods (no controls added)



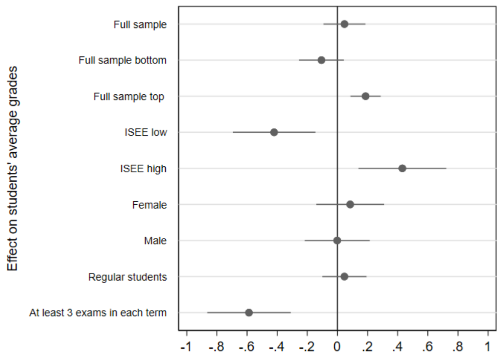

(a) Indicator for overall satisfaction (b) Indicator for salary increase

*Notes:* Estimation on academic years 2015/2016 - 2018/2019 (control group) and 2019/2020 (treatment group). Estimates of effects on SET for multiple time periods and by departments. Point estimates with 95% confidence intervals calculated based on standard errors clustered at the course level. Table A.2 shows the corresponding number of observations.

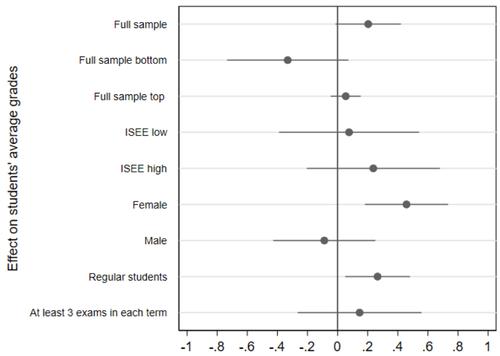
Figure C.9: Effects on average grades of online teaching due to COVID-19 pandemic, mandatory courses (no controls added)




(a) Full sample



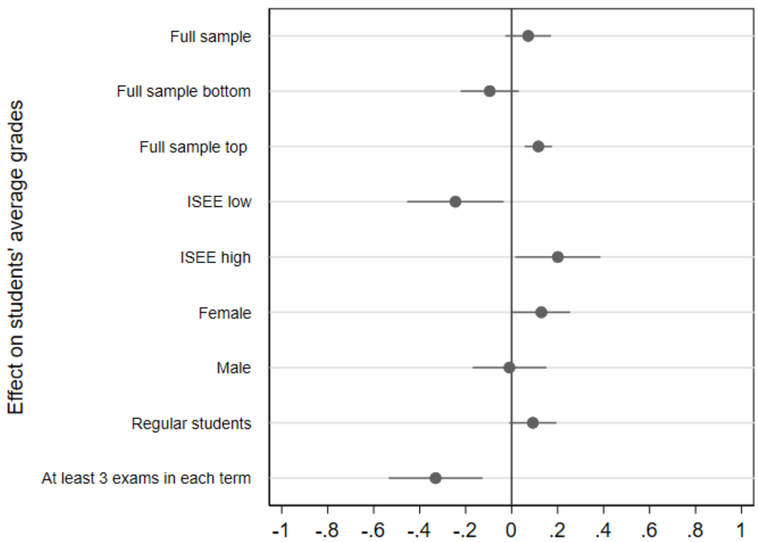

(b) Bachelor



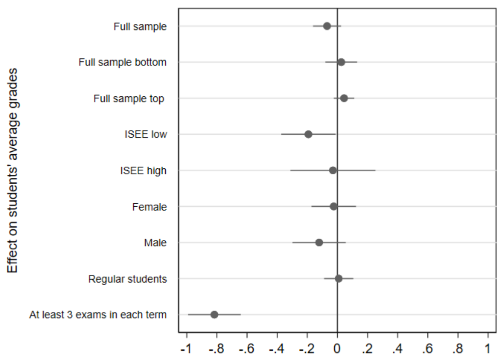

(c) Master

*Notes:* Estimates of effects on average grade. Full sample refers to mandatory courses. The figure shows estimates of the effects on students' average grade from a difference-in-differences specification without controls. Point estimates with 95% confidence intervals calculated based on standard errors clustered at the individual level. Table A.3 shows the corresponding number of observations.

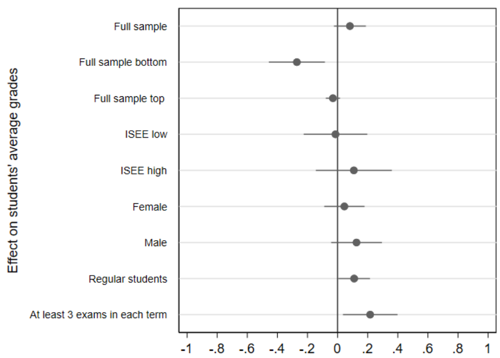
Figure C.11: Effects on average grades of online teaching due to COVID-19 pandemic, multiple pre-treatment periods (no controls added)




(a) Full sample




(b) Bachelor




(c) Master

*Notes:* Estimates of effects on average grade. Estimation on academic years 2015/2016 - 2018/2019 (control group) and 2019/2020 (treatment group). Full sample refers to academic years 2015/2016 - 2019/2020. The figure shows estimates of the effects on students' average grade from a difference-in-differences specification without controls. Point estimates with 95% confidence intervals calculated based on standard errors clustered at th individual level. Table A.3 shows the corresponding number of observations.