Institut d'Economia de Barcelona

IEB Working Paper 2024/03

MUDDYING THE WATERS: HOW GRADE DISTRIBUTIONS CHANGE WHEN UNIVERSITY EXAMS GO ONLINE

Zelda Brutti, Daniel Montolio

Version April 2024

Public Policies

IEB Working Paper

IEB Working Paper 2024/03

MUDDYING THE WATERS: HOW GRADE DISTRIBUTIONS CHANGE WHEN UNIVERSITY EXAMS GO ONLINE

Zelda Brutti, Daniel Montolio

The **Barcelona Institute of Economics (IEB)** is a research centre at the University of Barcelona (UB) which specializes in the field of applied economics. The IEB is a foundation funded by the following institutions: "La Caixa" Foundation, Saba, the Barcelona City Hall, the Barcelona Metropolitan Area, the University of Barcelona, the Autonomous University of Barcelona, the Barcelona Provincial Council, Agbar, Cuatrecasas and Consorci Zona Franca Barcelona.

The **IEB** research program in **Public Policies** aims at promoting research related to the design, implementation and evaluation of public policies that meet social needs and enhance social and individual welfare. Special emphasis is put on applied research and on work that tries to shed light on the Economics of Education, Health Economics, Innovation, Labour Markets and Security Policies. Disseminating research findings in these topics to a broader audience is also an aim of the program.

Postal Address:
Institut d'Economia de Barcelona
Facultat d'Economia i Empresa
Universitat de Barcelona
C/ John M. Keynes, 1-11
(08034) Barcelona, Spain
Tel.: + 34 93 403 46 46
ieb@ub.edu
http://www.ieb.ub.edu

The IEB working papers represent ongoing research that is circulated to encourage discussion and has not undergone a peer review process. Any opinions expressed here are those of the author(s) and not those of IEB.

MUDDYING THE WATERS: HOW GRADE DISTRIBUTIONS CHANGE WHEN UNIVERSITY EXAMS GO ONLINE *

Zelda Brutti, Daniel Montolio

ABSTRACT: We analyse how grade distributions change when higher education evaluations transition online and disentangle the mechanisms that help to explain the change observed in students' results. We leverage administrative panel data, survey data and data on course plans from a large undergraduate degree at the University of Barcelona. We show that grade averages increase and their dispersion reduce. Changes are driven by students from the lower end of the performance distribution and by a reduction in the occurrence of fail grades; however, we do not find evidence for artificial `grade adjusting' to explain the phenomenon. We are also able to dismiss shifts in the composition of test takers, improvements in teaching quality or in learning experiences and increases in student engagement. While changes in the assessment formats employed do not appear to mediate the causal relationship between online evaluation and higher grades, we identify more dispersed evaluation opportunities and increased cheating as explanatory factors.

JEL Codes: I23
Keywords: Higher Education; Online Education; Online Assessment; Administrative Data; Survey Data; Covid

Zelda Brutti
Joint Research Centre of the European Commission, Unit S.3 - CC-ME
E-mail: Zelda.BRUTTI@ec.europa.eu

Daniel Montolio
Universitat de Barcelona & IEB
E-mail: montolio@ub.edu

*We are grateful to Daniel Favre de la Noguera and Carles Pare Ogg for their valuable research assistance, to Marina Solé (Director of Studies of the University of Barcelona degree in Public Management and Administration in 2020) and Enriqueta Exp_osito (Vice-dean of the Faculty of Law of the University of Barcelona in 2020) for their collaboration to run the student survey during the pandemic, and to participants to the 2023 Modena workshop on "Higher Education and Equality of Opportunities", the 2023 AIEL Conference and the 2024 LESE Conference, as well as to two anonymous reviewers, for their helpful comments and insights. We acknowledge funding from the Spanish Ministry of Science, Innovation and Universities under the project PID2022-140661OB-I00; and from the project DEMOCRAT funded by the European Union's HORIZON-RIA HORIZON Research and Innovation Actions under Grant Agreement No. 101095106. Usual disclaimer applies.

1. Introduction

Online education has experienced an unprecedented surge in popularity in recent years, transforming the traditional learning landscape worldwide, thanks to the rapid proliferation of both the Internet and of technologies such as smartphones, tablets, and laptops. Significant improvements in the cost-effectiveness and in the accessibility of educational resources to individuals from diverse backgrounds and locations have been listed as beneficial outcomes of the online trend (Means et al. 2010; Coleman and Berge 2018); in higher education, students are provided with the freedom to customize their learning schedules, thereby accommodating their diverse lifestyles and commitments. Nevertheless, online learning has also been associated with a number of less desirable repercussions, such as inequality in outcomes due to gaps in technology access (Devkota 2021), larger room for cheating behaviours (Bilen and Matros 2021) and employers sometimes attributing lower value to online degrees with respect to traditional ones (Rosendale 2017). The evidence on overall impacts of online higher education on student outcomes is mixed, with some studies finding no significant difference in learning outcomes between online and in-person instruction (Bowen, Chingos and Lac 2014; Means et al. 2010) and others finding online instruction to be in general less effective (Xu and Smith Jaggars 2013).

The COVID-19 pandemic (Covid or pandemic, hereafter) further boosted the adoption of online education worldwide, as lockdowns and social distancing measures compelled educational institutions to transition rapidly to remote learning or hybrid modalities. This forced acceleration of the online format has underscored its benefits and its potential: by 2020, only a minority of individuals believed that higher education will be delivered mostly in-person in the near future (Whiting 2020)[1]. As online learning continues to evolve and innovate, it is poised to play an increasingly important role in shaping the future of education, calling for further research in this domain.

One question that has gone largely unaddressed by current research is whether and how grades obtained by students in higher education evaluations change when these are held online, as compared to the traditional in-classroom format. In this paper, we fill this gap in the literature by providing a systematic analysis of the reaction of undergraduate grade distributions as a result of the transition to online examinations.

As a case study, we use a large undergraduate degree at the University of Barcelona, which we observe transitioning online due to the Covid and on which we hold a valuable combination of administrative panel data, survey data and digitised data on course plans (syllabus). While recognizing that the pandemic represented an unusually large and unexpected shock to the educational environment and that it triggered other changes beyond the online transition, we believe that our analysis –thanks to the unusually rich wealth of information it leverages– is able to give a novel contribution to the literature. It not only gives a precise picture of how aggregate grade dynamics changed after online evaluation was introduced, but also sheds light on the heterogeneity in changes and on the mechanisms driving such changes, making a more convincing case for the online transition in particular to be the main causing factor. Additionally,

[1] Also sources of the European Commission confirm that interest in online education has been growing in the EU for a long time (`https://ec.europa.eu/eurostat/web/products-eurostat-news/-/edn-20220124-1`). Compared with 2019 before the pandemic, the share of people (16-74) doing online courses or using online learning material increased in all Member States, to an average of 27% in 2021 (23% in 2020). Being young people aged 16-24 those that take more online learning. In the US, the percentage of students enrolled exclusively on distance-learning courses rose from 12.8% in 2013 to 17.3% in 2019, according to data from the National Center for Education Statistics.

the fact that we observe one last academic semester combining online evaluations and traditional-format lectures further enhances our ability to disentangle the effects of online instruction from those of online examinations in particular.

As pointed out by Rodríguez-Planas (2022a), grades ensuing from an online evaluation setting in higher education can differ for various reasons. Students might be differently motivated or academically prepared with respect to the traditional framework; assessment processes may be different in terms of format or level of challenge; grading strategies may be different in terms of leniency; the frequency or intensity of cheating may change. All of these explanations may result in variations in the grades administered to students, but each of them has very different implications and yields very heterogeneous policy advice.

We show that when courses went online, grade averages increased and their dispersion reduced. This was driven by a reduction in failing grades and went in favour of students from the lower end of the performance distribution. We also find a relative performance drop of good students from lower socioeconomic backgrounds with respect to their peers and the opening up of a performance gap relating to the digital divide. Our systematic analysis of the potential drivers behind these results leads us to dismiss improvements in teaching quality, in learning processes, in student motivation and preparation and shifts in the composition of test takers during the online exam sessions. In terms of examination and grading processes, we are not able to cast judgments on the overall content of the exams or on their level of challenge, but we do find evidence against a systematic 'adjustment' of grades across the passing threshold. We find exam formats have changed, but in a way that is not statistically related to the change in results observed. The channels that we identify as likely contributors to the observed grade boosts and to the relative advantage of traditionally lower-performing students during online evaluations are assessment being less concentrated and cheating opportunities being more readily available in online assessment environments.

Grades obtained in university examinations are an important object of study, given that they influence students' behaviour and choices throughout the duration of the degree. Numerical exam grades obtained as well as the relative ranking compared to peers have been found to be a co-determinant of dropout behaviour from college majors (Astorne-Figari and Speer 2019) and STEM majors in particular (Rask 2010); both absolute and relative grades are well established in education and psychology literature as factors that impact motivation, perceived academic control, goal engagement and, ultimately, retention and graduation (Robbins et al. 2009; Hamm et al. 2019; Respondek et al. 2020). Last but not least, grade point average (GPA) becomes a key summary measure at the time of graduation and is used by students to signal their skills to potential employers upon entry to the job market (Hansen, Hvidman and Sievertsen 2024; Piopiunik et al. 2020). For these reasons, it is a necessary and worthwhile endeavour to investigate how and why individual grades and the relative performance of different student types change when evaluation is shifted online.

The rest of this paper is organised as follows. Section 2 gives a brief overview of literature related to our focus; section 3 describes the institutional framework; section 4 describes the data sources that we draw upon; section 5 presents the main results in terms of grade dynamics and heterogeneity between student types; section 6 shows our investigation of channels based on administrative data, while section 7 addresses results based on survey data; section 8 tests the explanatory power of different candidate channels; and, finally, section 9 concludes providing a policy-oriented discussion of results.

2. Closely related literature

There has been a substantial amount of studies that have investigated the consequences of the recent Covid pandemic on education and that have quantified the learning losses it created (see Patrinos, Vegas and Carter-Rau 2022 for a review of recent literature), for school-aged children and for specific demographic groups in particular. However, considerably less attention has been devoted to the exact mechanisms behind such results, to the role of online learning in specific, and to the population of higher education students.

Closely related to our research, there are a few case studies based on administrative data from US colleges. Rodríguez-Planas (2022a) finds that that the 'flexible grading' policy, which was implemented along with the online transition at Queens College in the City University of New York (CUNY), has helped lower income students to better deal with academic challenges. Bird, Castleman and Lohner (2022) use data from the Virginia community college system and a within-instructor-by-course variation on whether students started their spring 2020 courses in person or online, finding a negative impact of online learning on course completion. Orlov et al. (2021) analyse data from intermediate-level economics courses at Cornell University, in which student learning was measured by using standard multiple-choice assessments, and find a negative impact of the change to remote teaching due to the pandemic. Kofoed et al. (2021) randomised 551 West Point students between online and in person compulsory economics courses to find a negative impact on results by those students that took the course in an online format, a result mainly driven by below median academic ability. Tillinghast, Mjelde and Yeritsyan (2023) detect significant grade inflation among Texas A&M University undergraduates in pandemic-driven online sessions, even after accounting for institutional, instructor, and student characteristics, also finding that the relationship between some of these variables and GPAs changed during pandemic times. There is also evidence of grade inflation in universities outside the US as activities were shifted online, such as Karadag (2021) on Turkey and Toro (2022) on Colombia. Most of these studies do not carry out a thorough analysis of channels behind changes in grades, as we do in our study; moreover, most are unable to disentangle the effects of online instruction from those of online evaluation. In our paper we observe one academic semester in which instruction was carried out online but evaluation went back in its traditional format: since all relevant grade dynamics disappear in such circumstances, we are able to attribute most shifts to online evaluation in specific.

We also add evidence to studies that use student survey data, such as from Queens College (CUNY) (Rodríguez-Planas 2022b), from Arizona State University (Aucejo et al. 2020), or from the the first wave of the Global COVID-19 Student Survey that was administered between April and October 2020 (Jaeger et al. 2021). The online surveys employed in these studies explored the expected or materialised impacts brought about by the changes in the instructional setting on students' academic and labour market outcomes. Our survey was also administered in Spring 2020 and is centred on understanding the impact that the new academic setting had on students, distinguishing between a wide range of aspects, including their physical condition, mental well-being (Browning et al. 2021), and socio-economic issues.

Moreover, our survey includes questions regarding perceived cheating during online examinations (Bilen and Matros 2021).

3. Institutional Background

3.1 The online transition at the University of Barcelona

Face-to-face lectures and activities were suspended at Spanish universities on March 14th, 2020, as the first Covid lockdown was announced. Faculty were asked to provide a virtual alternative to traditional classes and to finish working through the material of ongoing courses; the exam sessions that took place at the end of that academic semester were all held in an online format.[2] While university management made significant effort to provide faculty and students with the necessary assistance in getting used to the new IT resources required for teaching and examining online, the sudden transition to distance learning was challenging for some of the academic population, especially among senior faculty members and the part of the student body lacking adequate IT resources at home (Vallespín 2020).

The University of Barcelona (UB) is a large, public university in Spain. As in most other public universities in the country, at UB the transition to the online format affected courses of the second semester of the academic year 2019/20, while the first semester of that academic year, including the exam session, had ended normally. Online classes continued over the first semester of the following academic year, 2020/21; in general, both faculty and students were more comfortable with the new arrangement by that time, so that fewer disruptions were recorded. Interestingly, our last semester of observation, the second semester of the year 2020/21, kept featuring online lecturing but transferred examinations back to the traditional in-person format. Table 1 summarises the timeline and format of the academic semesters and exam sessions around the Covid pandemic at the University of Barcelona.

Contrary to what occurred in some other universities in the US (Rodríguez-Planas 2022a), the University of Barcelona did not change its grading system in response to the Covid pandemic: without any pass/fail option for students to choose, course grades kept being given in numerical format on a scale from 1 to 10, with 5 being the threshold for passing the course.

3.2 The GAP degree at the University of Barcelona

The degree in Public Management and Administration, GAP degree hereafter,[3] is an undergraduate degree program whose normal duration is 4 years and that prepares students mainly for managerial careers in the public sector, such as local, national, or European institutions; the program's focal subjects are law, administration practices, human resource management, as well as public finance management. Vacancies on the GAP program are assigned through a centralized university admission system: students list their preferences for programs and are admitted based on their order of preferences and their university access grade, which is obtained on a standardised test at the end of high school, held in July of each year (PAU – *Pruebas de Acceso a la Universidad*). That is, degree programs admit their new students

[2] The academic year in Spain starts in September and ends in June. Universities usually can organise themselves into either three terms or into two semesters. In the case of semesters, the first runs from September to January, while the second semester from February to June/July.

[3] Acronym from the Catalan name *Gestió i Administració Pública*.

Table 1: Academic semesters and exam sessions around the Covid pandemic at the University of Barcelona

Academic Semester	Lectures	Assessment
2019-2020 Semester 1	Sep – Dec 2019	Jan - Feb 2020
(2019 S1)	Face-to-Face	In-Classroom
2019-2020 Semester 2	Feb – May 2020	Jun - Jul 2020
(2019 S2)	Face-to-Face until Mar 13th Online afterwards	Online
2020-2021 Semester 1	Sep – Dec 2020	Jan - Feb 2021
(2020 S1)	Online	Online
2020-2021 Semester 2	Feb – May 2021	Jun - Jul 2021
(2020 S2)	Online	In-Classroom

in descending university access grade order, until capacity is filled; a new 'minimum access grade' is determined each year, as a result of this process: it is the lowest grade that permitted access to that degree program, in that university, in that year. Depending on the exact academic year considered, there are between 80 and 120 vacancies available for new enrolments in GAP at the University of Barcelona and the program has become more competitive over time – as demonstrated by a rapidly increasing minimum access grade over the past decade.[4]

4. Data

4.1 Main sample

Our main sample is drawn from the administrative records of the GAP degree program at the University of Barcelona. We obtained data on GAP students enrolled during the academic years 2015/16 up until 2020/21. Students are identified through a unique student ID and the database offers information on basic demographics, first year of enrolment, university access grade, subjects enrolled to each year and semester, exams taken in each session, and their corresponding grades. The original sample was composed of 961 students, each observed at least once in any of the semesters between 2015 S1 and 2020 S2; each observation in the database describes a subject j to which a student i is enrolled in academic semester t.

A few sample restrictions have been applied in order to work with a group of observations that is not affected by outliers or atypical circumstances. We dropped observations referring to subjects that are either very recent additions to the GAP program or that were discontinued during the time-span 2015-2020; this removes around 8% of the original sample. We further dropped observations that refer to subjects that are not consistently taught in the same academic semester over the period of

[4]The minimum access grade for the GAP program at the University of Barcelona was 5.00 in the academic year 2017/18 and grew steadily to reach 7.12 in the academic year 2020/21. Source: https://universitats.gencat.cat/ca/detalls/oferta/1193-Gestio-i-Administracio-Publica,lastaccessedonJuly2023.

observation; this removes around 13% of the remaining observations. Finally, we drop observations that refer to subject-course combinations with fewer than 25 students enrolled; this cuts approximately the lowest 5% tail of the remaining observations.

Our final sample of analysis consists of 766 individual students who were followed over time spans of varying length during the 2015-2020 window. Table A.1 in the Appendix shows descriptive statistics for this sample of individuals. Our average student is somewhat more likely to be female (55% vs 45% male), is 21.5 years old, has an individual grade point average of 5.85, does not show up to exams 2.3% of the time, entered the program with a university access grade of 6.2 and has a 35% probability of benefiting from a scholarship or from financial help.

4.2 Survey sample

During May-June 2020, and in collaboration with the Office of the Dean of the Faculty of Law and the Director of Studies of GAP, we designed and implemented a student survey to understand the impact that the new academic circumstances were having on students. The survey was answered at the end of the teaching period; approximately half of respondents answered the survey during the period in which exams were taking place.[5] The survey was set as available to all students via a link at the Virtual Campus of the GAP Degree. Communication with students regarding the survey was carried out by the Director of Studies of the GAP program and several reminders were sent to students.

The survey elicited information on physical, mental and family socio-economic conditions (using Likert scales) both before and during the online period, in order to allow for comparisons over time.[6] Other sets of questions addressed student engagement in academic activities, their access to digital resources, their perception of the online teaching and learning environment, and of cheating among students during evaluations.

The survey was answered by 98 individuals, of which 93 survive the sample restrictions previously described. A priori, we would not expect survey respondents to be a representative subset of the main population described in the previous section. To start with, given that the survey was administered in June 2020, the students surveyed were necessarily drawn from the population of students still in the process of completing their degree during the academic year 2019/20: we call this the group of 'potential survey respondents'. Most of the students we observe at *any* point during the 2015-2020 period are not potential survey respondents, since they finished or dropped off the program before the academic year 2019/20 or, in a smaller fraction, started it in 2020/21. In fact, the group of potential survey respondents is composed of 331 individuals, who we observe being enrolled in at least one subject in 2019/20; their descriptive statistics are summarised in Table A.1 in the Appendix. On average, potential survey respondents are 'better' students with respect to the main population, as suggested by higher GPAs, slightly higher university access grades, lower shares of no-shows at exams and higher probability

[5] The exam period started Monday, June 8th and extended until Wednesday, June 23rd. 47.5% of our respondents fulfill the survey between June 5th and June 7th, and the proportion grows to 64.6% by June 14th. We do not pursue any specific analysis on the timing of survey response, and assume that responding after having sat any online exam simply reduces uncertainty about one own's online performance.

[6] As example, and regarding physical condition Question 11 was *"On a scale from 1 (very poor) to 10 (very good), how would you define your physical health PRIOR to COVID-19?"*, while Question 12 was *"On a scale from 1 (very poor) to 10 (very good), how would you define your physical health DURING COVID-19?"*.

of having earned a scholarship. Many of these differences are explained by the fact that the GAP program at the University of Barcelona has been improving the quality of its student intake over the years, as mentioned previously in Section 3. Finally, the 93 survey respondents described in Table A.1 and representing 28% of potential respondents, are further positively selected in terms of GPA, access grade, share of scholarships and exam no-shows. This is in line with the intuition that those students who decide to answer a survey of the kind we administered for this analysis are likely to be part of the most diligent and academically involved group; the survey sample is also more female-biased with respect to the population of origin.

Although we are aware that our respondents do not constitute a representative sample of the student population enrolled in the GAP program at the University of Barcelona, we are aware of the kind of selection that we are facing; furthermore, our empirical analysis shows that despite the presence of selection, many of the most important grade dynamics ensuing from the transition to the online setting are similar between survey respondents and the larger samples from which they are drawn. We, thus, expect our survey data to help us in shedding light on the mechanisms behind the changes in grade patterns detected in the main sample.

5. Analysis of grade dynamics

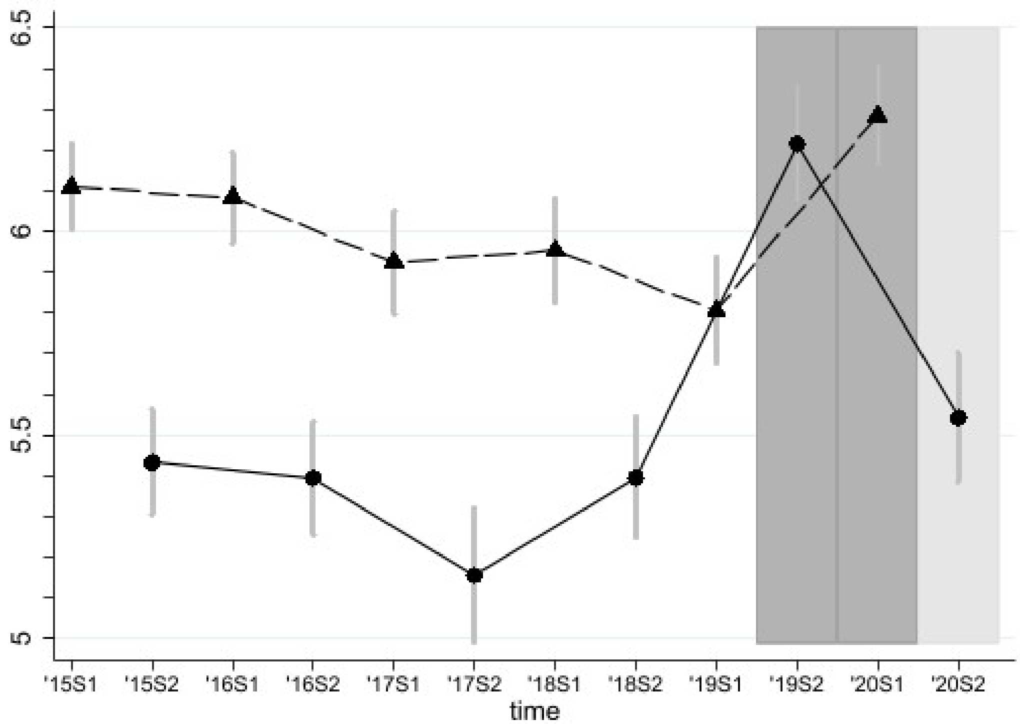
We begin our empirical analysis by describing the evolution of exam grades in our main sample of observation. Throughout most of the empirical analysis, we look at average grades separately by semester, given that the two academic semesters display a large and systematic difference in grade averages, with first semester grades being higher than second semester grades. This is explained by the different subject composition of the two semesters: on average, GAP students perform better in the first-semester (S1) than in second-semester (S2) subject mix. We call an ‘exam session’ a combination of academic course and academic semester: that is, the exam sessions that we are able to observe in the data are 2015 S1, 2015 S2, 2016 S1 and so forth, until 2020 S2; in each session, students sit exams of several different subjects, according to their degree plan and individual situation on the academic path.

Figure 1 highlights our first empirical finding: average grades increased noticeably during online sessions, particularly in the first affected session (2019 S2), but also on the second affected session (2020 S1). During both of these sessions, shaded dark grey in the graph, examinations were held in an online format, as explained previously in Section 3.1. Strikingly, grade averages fall back into their normal range in the 2020 S2 session, shaded light grey, where examinations went back to their traditional in-classroom format, although the corresponding lectures were still being held online.

A quantification of the change in grade averages along the different academic semesters can be obtained through a linear regression exercise. Table 2 shows the results of regressing standardised exam grades on academic semester-fixed effects as well as on individual student-level characteristics. Results in the table are expressed in pre-pandemic, exam-level standard deviations and the reference academic year is 2015.

The results of the quantification exercise confirm the evidence highlighted previously in Figure 1, that is, grade averages are significantly higher with respect to the norm in both academic semesters the assessment sessions for which took place online. The phenomenon is particularly pronounced in the first

Figure 1: Evolution of mean grades in the main sample. 1st semester dashed, 2nd semester solid



assessment session held online, 2019 S2, in which grade averages outstrip the typical second semester levels, which had been quite stable over time, by almost 50% of a standard deviation. First-semester grade averages are generally higher than second-semester averages and had been displaying a decreasing pattern over the years. The second assessment session held online, 2020 S1, saw an increase of almost 10% of a standard deviation with respect to its reference semester, or of about 20% of a standard deviation with respect to the previous year.

The regression setup also allows us to account for individual student characteristics when quantifying the average effect of online exam sessions. Model (2) augments the baseline regression by including student characteristics, in order to account for the fact that these might change over time, while Model (3) further adds subject fixed effects. The quantification of the online-session effects is not changed significantly by these additional features and the inclusion of subject effects, if anything, increases the estimated magnitude; therefore, given the specification in Model (2) is more parsimonious and slightly more conservative, we take it as our preferred. Figure 2 provides a graphical illustration of the academic semester effects estimated in Table 2. Note that regarding this first result, the occurrence of grade inflation in higher education programs during the Covid pandemic, has been documented in previous studies on Turkey (Karadag 2021), Texas (Tillinghast, Mjelde and Yeritsyan 2023) and Colombia (Toro 2022).

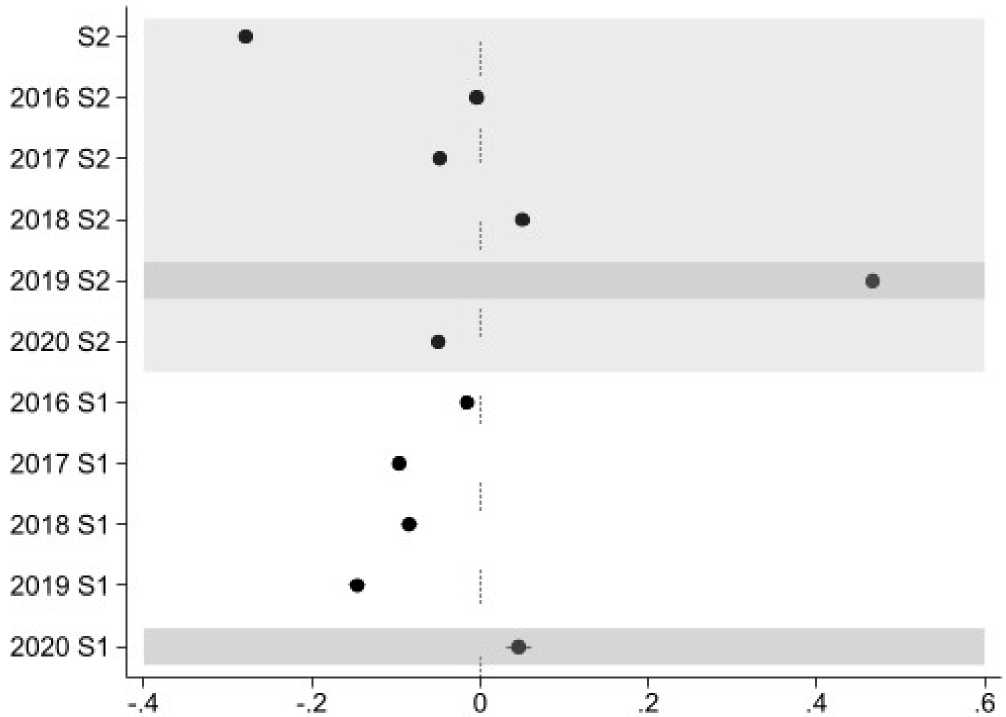
We expand our descriptive analysis of changes in grade dynamics by focusing on grade dispersion. Figure 3 describes the evolution of grade standard deviations over our period of observation: each marker represents the standard deviation of the grades obtained in a specific evaluation session by students belonging to our main sample. Grade dispersion is typically higher in second-semester sessions with

Table 2: Impact of online exam sessions on grade averages

	(1) Baseline	(2) Individual controls	(3) Subject FEs
2016 S1	-0.01***	-0.02***	-0.04***
	(0.00)	(0.00)	(0.01)
2017 S1	-0.08***	-0.10***	-0.13***
	(0.00)	(0.00)	(0.00)
2018 S1	-0.07***	-0.08***	-0.10***
	(0.00)	(0.00)	(0.01)
2019 S1	-0.13***	-0.15***	-0.15***
	(0.00)	(0.00)	(0.01)
2020 S1	0.08***	0.05***	0.07***
	(0.00)	(0.01)	(0.01)
S2	-0.30***	-0.28***	0.18
	(0.00)	(0.00)	(0.18)
2016 S2	-0.00***	-0.00*	0.04***
	(0.00)	(0.00)	(0.01)
2017 S2	-0.04***	-0.05***	-0.00
	(0.00)	(0.00)	(0.01)
2018 S2	0.05***	0.05***	0.09***
	(0.00)	(0.00)	(0.01)
2019 S2	0.47***	0.47***	0.52***
	(0.00)	(0.01)	(0.02)
2020 S2	-0.03***	-0.05***	-0.07***
	(0.00)	(0.00)	(0.02)
Age		0.01***	0.00
		(0.00)	(0.00)
Female		0.08**	0.07**
		(0.03)	(0.03)
Spanish		0.30***	0.26***
		(0.06)	(0.07)
Access grade		0.21***	0.20***
		(0.02)	(0.01)
Scholarship		0.26***	0.23***
		(0.04)	(0.03)
Subject FEs	No	No	Yes
Mean(y)	0.03	0.03	0.03
sd(y)	1.00	1.00	1.00
N.obs	13,119	13,072	13,072
R-squared	0.02	0.09	0.19

Notes: Linear regression results of exam grades on academic semester fixed effects and a set of student observables and subject fixed effects. The dependent variable is individual exam grades expressed in pre-Covid, exam-level standard deviations in the main sample. 2015 S1 is the reference group. SEs clustered at the year-semester level are in parentheses. p < 0.10 *, < 0.05 **, < 0.01 ***.

Figure 2: Estimation of academic semester fixed effects, holding constant student observables.

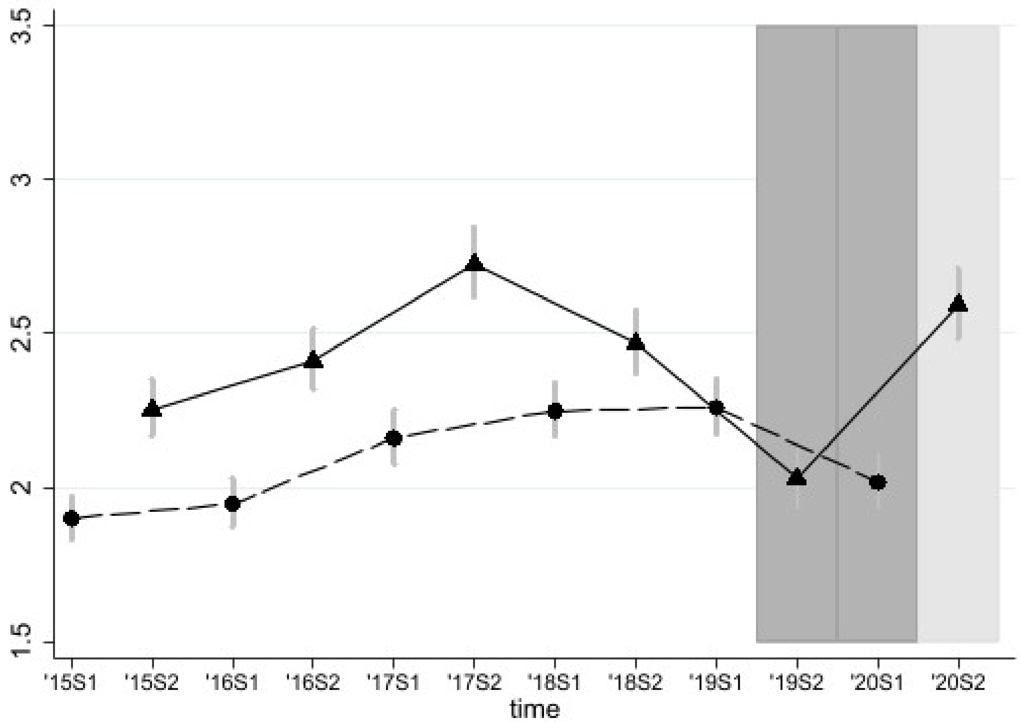


respect to first-semester sessions. Similarly to what was observed for grade averages, the two academic semesters in which evaluation were held online, shaded in dark grey, stand out with respect to the norm: a marked drop in grade dispersion can be appreciated for both semesters, and again the phenomenon is particularly strong in the first of the two sessions held online, 2019 S2, in which the grade standard deviation dropped even below the values that are typical for S1 exam sessions. Dispersion remains lower-than-usual in 2020 S1 and then bounces back to its historical levels in the 2020 S2 session, whose corresponding lectures were still held online, but evaluation for which went back into the classroom.

A more detailed appreciation of the change in grade distributions, when switching to online assessment, can be drawn from Figure 4. The two panels show the full distribution of online session grades (solid line) compared to the preceding 'normal' session (dashed line); the panel on the left refers to S2 sessions and the one on the right to S1 sessions. We can observe that the increase in the average and the contemporaneous reduction in dispersion which was identified previously in this section are produced by a reduction in the mass of grades below 5, the passing threshold. In 2019 S2, in the first online session, we can observe that the mass 'missing' in low grades, as compared to the previous comparable assessment session, is distributed in values slightly above 5 as well as at higher values between 8 and 10. In 2020 S1, we also observe a 'missing' mass of fail-grades, which appears to distribute smoothly across all values above the passing threshold. When exploring the possible factors driving the change in grades, Section 6.2.1 investigates the question of a potential 'artificial' manipulation of results by evaluators in greater depth.

The general message that may be taken away from this first, descriptive part of the analysis is that the two evaluation sessions that were held online displayed pronounced differences in terms of

Figure 3: Evolution of grade standard deviations in the main sample. 1st semester dashed, 2nd semester solid



examination results with respect to the historical patterns characterizing the GAP degree at the University of Barcelona. Grades given to students were on average higher, as a result of a reduction in the proportion of fail grades. In the first and unanticipated online session, 2019 S2, the empirical evidence points towards a shift from failing to mainly just-pass and pass-grades. The shift away from fail grades persisted in the following online session, 2020 S1, smoothly increasing the frequency of results above the passing threshold. Overall, these shifts have resulted in a less dispersed grade distribution with a higher mean value.

5.1 Heterogeneity analysis

5.1.1 Student quality

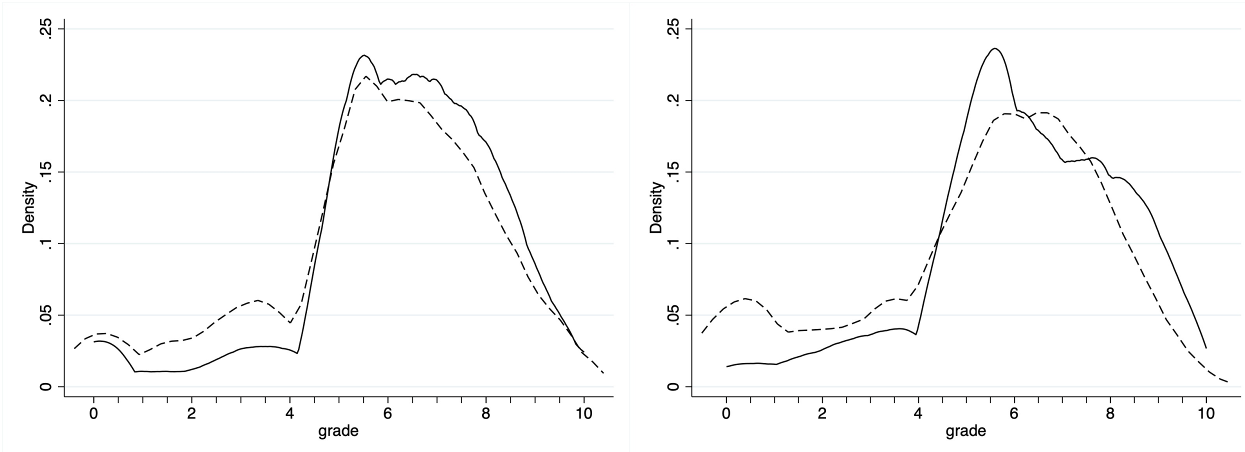
Having established that online evaluation sessions witnessed an increase in average grades, this subsections and the ones that follow focuses on how the change distributes across different types of students. Beyond pinpointing the main beneficiaries of the grade increase more precisely, the heterogeneity analysis is instrumental towards gathering insights about the mechanisms behind the aggregate results observed.

The first dimension of heterogeneity that we explore is student quality, in order to establish whether the increase in performance during online sessions is homogeneous across different levels of student performance. In order to do so, we use individual pre-Covid GPAs to classify students of our main sample into three performance groups: high-performing students are those whose pre-Covid GPA was above the '*notable*' (in Spanish grade terminology) threshold of 7 (19%); average-performing students

Figure 4: Grade distributions in the main sample, traditional (dashed) vs. online (solid) semesters

(a) 1st semester

(b) 2nd semester

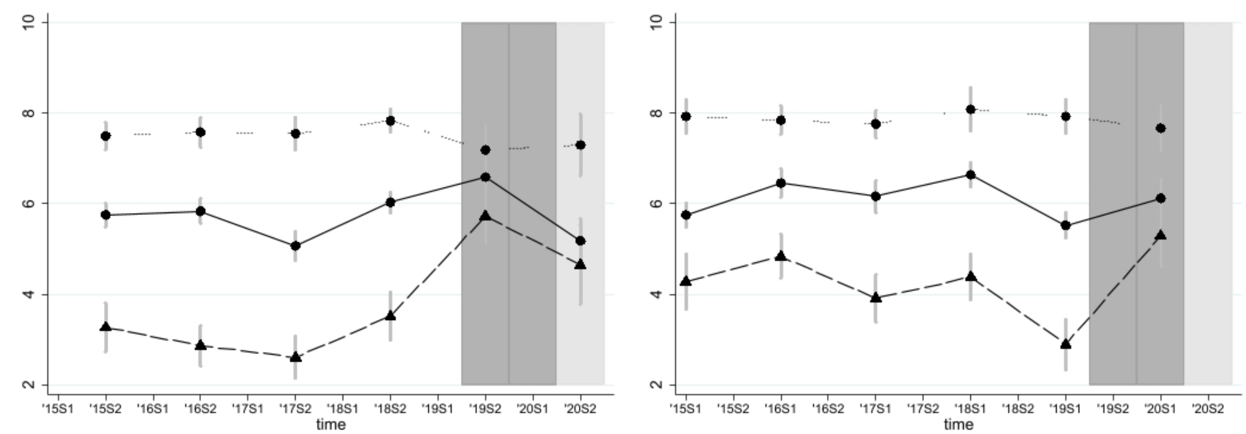


are those whose pre-Covid GPA was above the ‘pass’ threshold of 5 but below 7 (46.5%); low-performing students are those whose pre-Covid GPA was below 5 (34.5%). Figure 5 plots the evolution of average grades for each of the pre-Covid performance groups.

Figure 5: Evolution of mean grades in the main sample, by pre-Covid GPA groups

(a) 2nd semester

(b) 1st semester

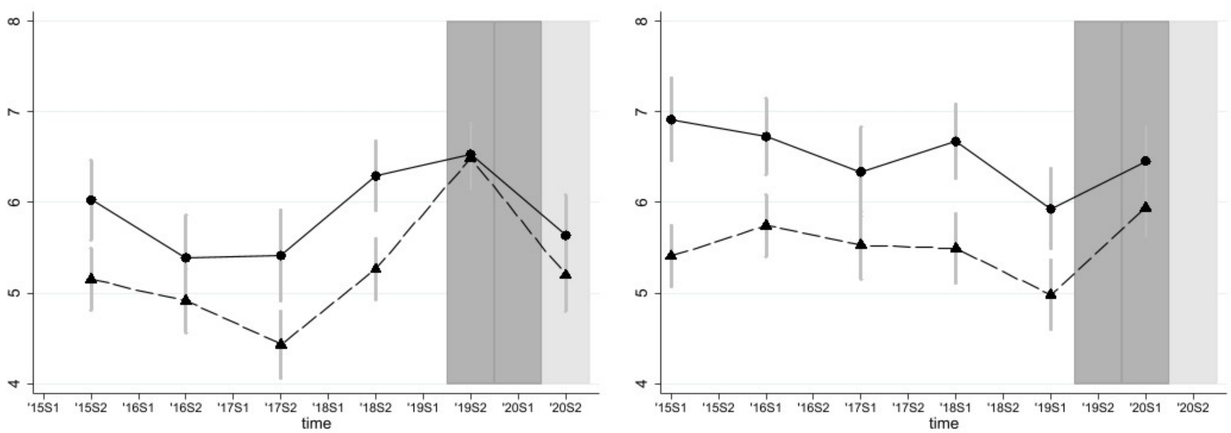


The insight we obtain from Figure 5 is that most of the aggregate dynamics that we had observed in the previous Figure 1 are driven by the group of low-performing students, who experience the most striking increase in mean grades during the online sessions 2019 S2 and 2020 S1. Average-performing students are shown to contribute to increasing aggregate average grades as well, albeit in a much lower fashion. Both of these two groups of students witness a drop in their grade average for 2020 S2, when teaching was still online but after examinations had gone back to the traditional in-classroom format. Finally, the average high-performing student did not see any benefit in terms of grades during online sessions; in fact, if anything, we can observe a drop in average GPA among these individuals as evaluation went online.

Figure 6 shows the result of a robustness exercise that uses a second measurement of student quality, thereby implying a simpler split of the sample. Students are split into two performance groups according to their university access grade: low-performance with a below-average access grade (below 6.20) and high-performance with an above-average access grade (above 6.20). The results from this additional analysis confirm our previous ones and are in line with similar findings from Turkey (Karadag 2021), that is, the largest online grade gains are concentrated in the lower-performing group of students; in Figure 6, a convergence between the two groups is visible as a result.

Figure 6: Evolution of mean grades by access grade group (high grade solid, low grade dashed)

(a) 2nd semester (b) 1st semester



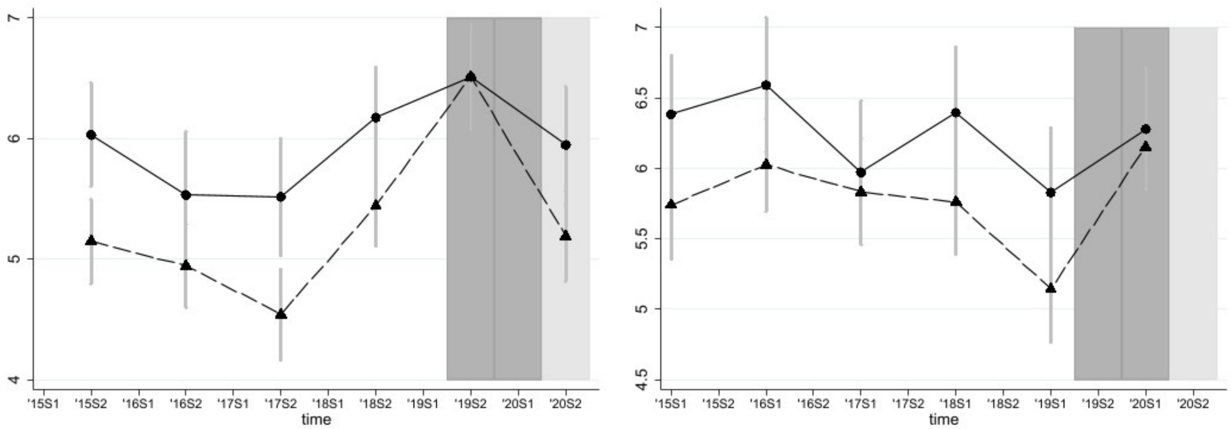
5.1.2 Socioeconomic status

A body of recent research has documented significant differences in the impact of the Covid pandemic and the consequent change in instruction and evaluation methods across socioeconomic groups, both at the level of compulsory schooling (Andrew et al. 2020; Barnum and Bryan 2020) and at the college level (Rodríguez-Planas 2022b and Aucejo et al. 2020). Most of the results point at stronger and more negative academic performance impacts for students from more disadvantaged backgrounds, as a result of the digital divide and the asymmetry in economic and health shocks.

A proxy of socioeconomic status offered by our administrative database is students' scholarship status. The vast majority of publicly-funded university scholarships in the Barcelona area are either solely based on income and have the explicit purpose of guaranteeing access to higher education to disadvantaged pupils (*Beques equitat* – 'Equity scholarships'), or are based on a combination of academic merit and income, thereby excluding pupils from families above given income and wealth levels (*Beques generals* – 'General scholarships'). As a consequence, public scholarship holders represent a sample that is negatively selected on socioeconomic background and positively selected on academic performance; as illustrated in Table A.1, about 30% of students in our main sample benefits from a scholarship. Figure 7 replicates our main descriptive results in terms of grade averages, distinguishing between scholarship holders (solid) and students without a scholarship (dashed). The main insight that this figure offers is that while scholarship holders outperform their peers in normal times, the gap closes strikingly during the special online Covid exam sessions, 2019 S2 and 2020 S1, just to open up again when exams returned to their traditional format in 2020 S2.

Figure 7: Evolution of mean grades by scholarship status (with scholarship solid, dashed without)

(a) 2nd semester (b) 1st semester



This result confirms the conclusions of Rodríguez-Planas (2022b), who finds that students from lower socioeconomic background in Queen's College (The City University of New York) achieved a performance advantage over their higher-income peers during pandemic sessions, but only because they were offered and more frequently made use of the pass/fail grade option – and that they would have suffered a relative performance loss if not offered such option. Based on students' survey responses, Rodríguez-Planas (2022b) concludes that such counterfactual disadvantages would be due to greater challenges with online learning and higher concerns about maintaining financial aid among disadvantaged students. As explained in Section 3.1, the University of Barcelona did not implement a pass/fail grade policy and we indeed obtain the expected relative performance loss by lower-income students with respect to their higher-income peers. In the second part of the paper, we will discuss some additional anecdotal evidence from survey data backing up this point.

5.1.3 Gender

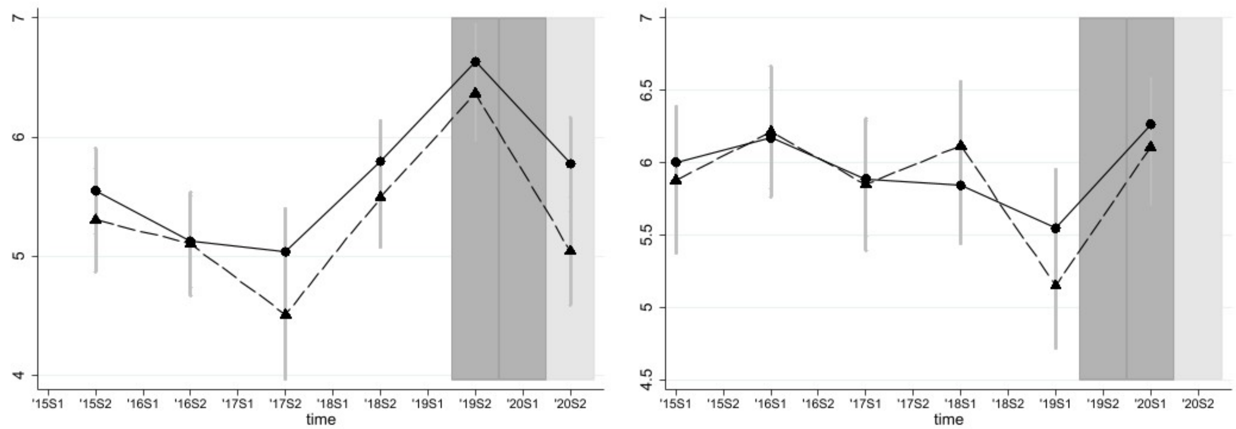
We asked ourselves the question of whether grade gains might be concentrated predominantly in either gender group – a conjecture which is backed up by recent literature (see Ayllón 2022). Thus, we divided our sample by gender and looked at the evolution of grade averages separately for the two groups. Results are shown in Figure 8 and can be summarised by stating we are unable to detect any sizeable gender differences in the average grade dynamics during online evaluation sessions in our sample.

5.1.4 Technology

As pointed out in previous studies (Devkota 2021), technological gaps are a significant dimension of heterogeneity to explore when assessing the impact of online instruction and evaluation. In a remote education context, students' preparation and exam results are likely to be influenced by the technology available in their homes. Like most other countries worldwide, Spain had recognised this issue early on at the onset of the pandemic and had deployed large amounts of financial resources earmarked for the reduction of technological inequalities among students (see *Programa Educa en Digital* and OECD 2021),

Figure 8: Evolution of mean grades. Male students (dashed) and female students (solid)

(a) 2nd semester (b) 1st semester



and individual universities themselves have further topped-up such funds where possible.[7] However, a digital divide may have persisted in the population of university students and we are able to explore the issue empirically. We do not possess information about technological endowment for the full student population, so that we reach out to our survey sample to address this dimension of heterogeneity; note that in Section 7 we will discuss in detail how our survey sample behaves similarly to the general population in terms of the observed graded dynamics. There is a specific survey question that addresses the issue of connectivity at home, and it allows us to divide respondents into a group of "well connected" and of "less-well connected" learners;[8] Figure 9 illustrates the evolution of average grades in the two groups. Even though precision is low due to the limited size of the survey sample, a performance gap appears to open between the two groups of students when lectures and examinations go online, in favour of the one endowed with high-speed connections at home.

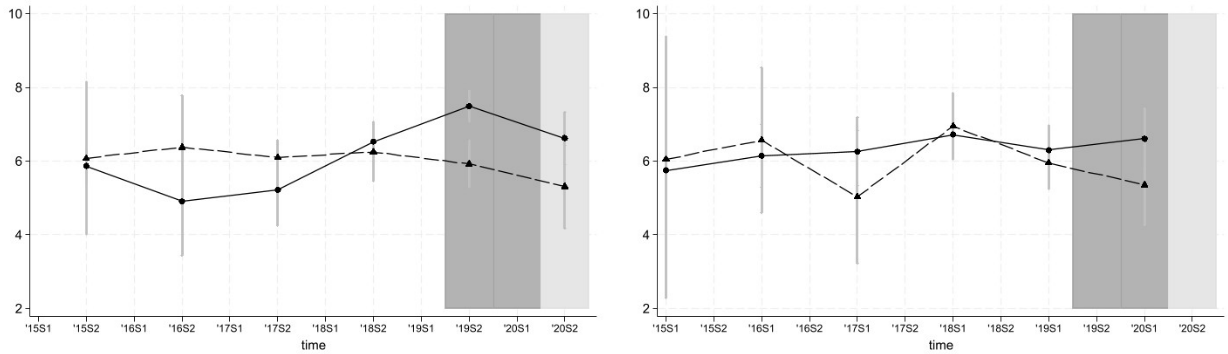
It is worth noting that, contrary to most of our other findings in this paper, the performance differences emerged during the online transition did not dissipate in the last semester of observation, 2020 S2, when lectures were still online but exams went back to their in-classroom format. This suggests that when it comes to learning technology, remote *evaluation* and *instruction* may cause *heterogeneous* deviations in university student outcomes with respect to the traditional setting. This finding supports those from previous literature, which have stressed the inequality-generating role of the digital divide when education transitions online (among others, see Correia 2020, Azubuike, Adegboye and Quadri 2021, Eruchalu et al. 2021, Ramsetty and Adams 2020). In this sense, we find that the institutional efforts towards a reduction of the gaps in access to adequate learning technology are warranted.

[7] At the University of Barcelona, the program CONNECTA-UB defined a number of initiatives aimed at ensuring Internet connectivity and providing temporary use of laptops to students.)

[8] The question posed to students is *What is your degree of connectivity for studying? 1) Where I study, I have access to high-speed internet connection (fibre optic). 2) Where I study, I have access to regular internet connection (ADSL). 3) Where I study, I have access to internet connection via my mobile data. 4) Where I study, I have access to internet connection via satellite. 5) I don't have access to internet connection.* based on responses to the connectivity question, we split the survey sample into two groups, the first declaring to have optic fibre or satellite connection at home (61.7%) and the second to have other type of slower connections (38.3%).

Figure 9: Evolution of mean grades among surveyed students, by available technology at home (high-speed internet connection solid, other type of connection dashed)

(a) 2nd semester (b) 1st semester



6. Drivers: evidence from administrative data

In this section, we begin our analysis by exploring the potential drivers behind the changes in grade dynamics observed during online evaluation sessions.

6.1 Compositional effects

One possible explanation for a change in grade results is a change in the composition of the pool of exam takers. During online evaluation sessions, the type of candidates may be different with respect to standard sessions. The new examination method and the higher uncertainty about outcomes ensuing from it might induce exam withdrawal and postponement by *low-performing students* fearing failure, or by *high-performing students* seeking to avoid damaging their GPA ambitions. However, the opposite sort of conjectures are likewise plausible: a larger-than-usual number of students might decide to attempt exams when these are offered online, driven by hopes of taking advantage of the online environment.

To explore compositional effects on the pool of exam takers, we compute the number of exam takers registered for each evaluation session, as well as the share of absences ("no-shows") in each of these. Many of the possible conjectures about changes in the type of exam candidates during online sessions also imply a change in the *overall number* of candidates ensuing as a consequence. The first panel of Figure B.3 in Appendix B plots the evolution in the number of exam takers during our period of observation; the main conclusion is that online sessions did not display any anomalies along that dimension. The same conclusion holds for the share of 'no-shows,' represented in the second panel of Figure B.3; that is, online sessions do not exhibit changes in the proportion of students who are due to sit exams, but who fail to do so. Finally, the bottom panel of Figure B.3 performs a similar check on the total number of university credits that students are enrolled for in each evaluation session. Credits are a proxy measure of the workload required by each subject and related exam, and one might speculate that students shift towards "heavier" or "lighter" exams during the exceptional online sessions. However, as the figure shows, online sessions do not stand out with respect to traditional ones along this dimension either. The evidence collected so far in this subsection can be summarised by saying that the descriptive data appears to exclude sizeable compositional changes in the pool of exam candidates during online sessions.

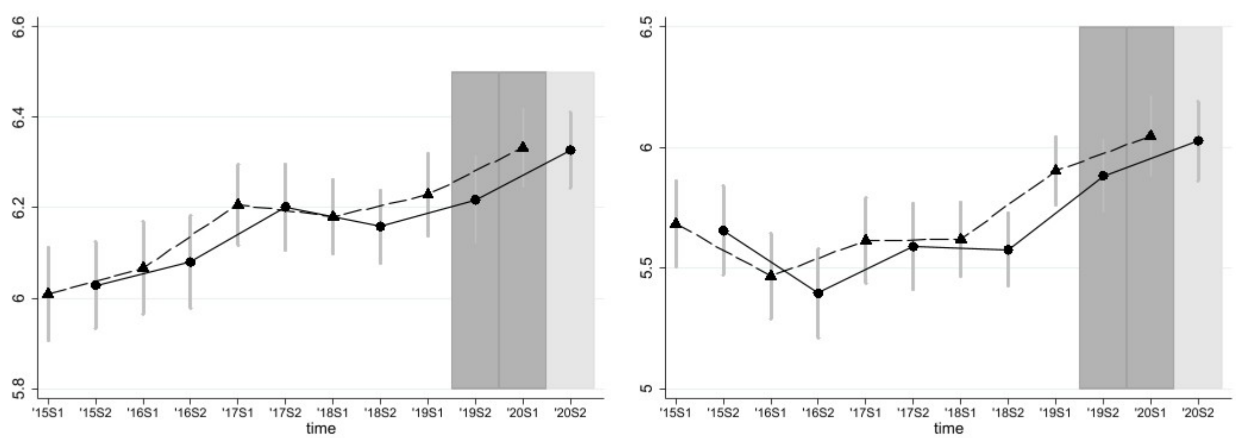
Comparing online exam sessions with those in traditional format, there are no appreciable differences in the number of individuals enrolled, in the share of exam 'no-shows,' or in the average number of credits taken up.

Despite the absence of evidence of sizeable quantitative shifts in exam session attendance, at this stage one cannot rule out completely a change in the *average quality* of candidates during online sessions – a phenomenon which might help to explain the aggregate grade dynamics described in the first part of the analysis.

Finding an ideal measure of the academic quality of exam takers in this setting is a nontrivial task and we present results using two different options. The first option is based on the university access grade of students taking exams in any specific session: as described in subsection 5.1.1, this grade represents an evaluation of students' academic ability at the end of high school; the first panel of Figure 10 plots the average university access grade calculated among the pool of exam candidates in each session. A necessary remark is that university access grades have been on a general rising trend in Spain for several years now (El País 2022), amidst a general discussion on grade inflation and a lowering of standards. In Figure 10, averages in the two semesters of any year are very similar, which is explained by the fact that the typical student takes exams in both semesters of a given academic year and thus contributes to both pools of candidates, S1 and S2; exceptions are given by students that are following a less-regular degree path because of delays, anticipations of specific exams, Erasmus experiences, and so on.

Figure 10: Average quality of the pool of exam takers; first semester dashed, second semester solid

(a) Average university access grade (b) Average pre-Covid GPA



On a side note, the graph shows that both semesters of the year 2020 witnessed a larger-than usual increase in access grades. In order to interpret this phenomenon, one must keep in mind the fact that Spanish university access examinations themselves were also affected by the Covid pandemic in their July 2020 and July 2021 sessions. Students were allowed a wider selection of topics to choose from when sitting the exams and a wider use of multiple-choice test formats, so that these extraordinary sessions were perceived as easier by students; the share of very high grades increased notably with respect to previous years (El Español 2022; El Correo Web 2022; Europa Press 2021). This phenomenon explains the increase we that observed in the last portion of the plot in Figure 10a, beyond the general upward trend.

If the new online-format for exams had induced a change in the quality of candidates, following strategic behaviour discussed previously in this section, then the change ought to be visible in the plot. What we observe instead is that the first and most troublesome online session, 2019 S2, saw a pool of candidates whose average access grade was very much in line with the previous, traditional-format session, 2019 S1, suggesting that roughly the same type of students enrolled for exams in those two sessions and confirming the previous evidence of no significant compositional changes in the pool of exam takers as sessions transferred online.

The second option for tracking compositional changes in the pool of exam takers is to use GPA measured in pre-Covid sessions as a proxy for student quality. The second panel of Figure 10 plots the evolution of the average pre-Covid GPA among candidates at each session. The downsides of this measure of candidate quality are in part similar to the one highlighted regarding the access grade –i.e., the pre-Covid GPA has a natural upward trend because students admitted to the GAP program are getting better over the years. Additionally, sample attrition plays a role in explaining the positive trend of this quality measure. Given that university dropout rates are high in Spain and at the University of Barcelona (Freixa Niella, Llanes Ordóñez and Venceslao Pueyo 2018) and given that better students tend to stay longer on the degree program, it is normal that conditioning on the existence of a pre-Covid characteristic such as grade GPA, students observed during Covid semesters and later would be a positively selected sample of the population. It is important to notice that the somewhat larger GPA growth observed during the last two years starts already in the first, traditional semester of 2019. Similarly to what we concluded about university access grades, an inspection of the GPA dynamics does not reveal any sudden change attributable to strategic selection into or out of the special online exam sessions of both 2019 S2 and 2020 S1.

6.2 Assessment-related factors

The increase in average grades observed when evaluations moved online may be the result of factors directly related to the way in which assessment is managed at the higher education institution. For instance, the academic quarters whose evaluations were moved online may have benefited from a more generous *grading protocol*, applied due to the exceptional circumstances. Alternatively, exams may have been *easier* with respect to the level usually administered in traditional evaluation sessions. It is nearly impossible to cast an accurate judgment on each of these aspects, due to the intrinsically subjective concept of "difficulty," as well as the large variability in factors such as course content, instructors, teaching methods, and assessment criteria. In this section, however, we provide some evidence that we were able to gather on empirically quantifiable aspects concerning grading and assessment types.

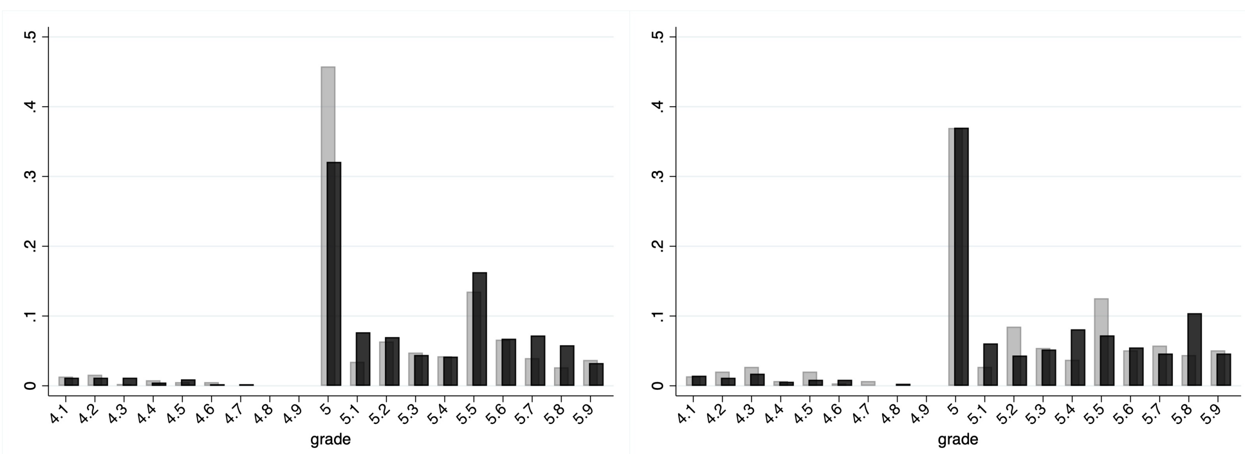
6.2.1 Grading

In terms of detecting possible biases towards more generous grades, we focus on one measurable phenomenon involving intentional 'grade rounding' around specific thresholds. More specifically, it is not unusual for evaluators to round the mark actually obtained by the candidate up to the following integer at the end of the marking process of an exam or course performance –especially if this means passing from a lower grade class to a higher (fail / pass / merit / distinction). Markers may do this to compensate for any mistakes they may have committed in judging the exam, as well as to 'help' candidates reach the subsequent grade class if they are very close thereto, since the grade classes matter for a student's curriculum beyond the numerical GPA. For the reasons just mentioned, rounding just-fail marks up to the passing threshold is especially common.

As we had already noticed in Figure 4, grade density plots show that online exam sessions display an excess mass of passing marks with respect to fail marks, when considering the sessions just pre-Covid as the baseline. Here we focus on grades around the passing threshold and explore the question of whether the excess mass detected is concentrated specifically at the threshold, that is 5, in order to spot any evidence of larger-than-usual mark adjusting practices. Figure 11 shows the detail of grade distributions from our administrative data, focusing on grades between 4 and 6. The 'rounding' practice is immediately visible when looking at the distributions, which all feature particularly low concentration of grades lower but close to 5 and high frequencies of candidates scoring exactly '5'. If graders applied a more lenient grading strategy, so to 'help' students achieving the pass mark, then we should be observing a higher concentration of '5' grades during the online sessions with respect to the traditional sessions. However, the pairwise comparison between the grade distributions two online sessions and their preceding, traditional-format counterparts shows no evidence of excess 'rounding' having been applied as exams went online. If anything, when looking at S1 results, the online session in 2020 displays significantly fewer 'exact 5' results with respect to the previous year, 2019. From this more detailed analysis of grade distributions around the pass/fail range, we can draw the conclusion that the passing marks surplus recorded during online session are spread over results above and not necessarily close to the '5' threshold, which makes artificial grade adjusting less likely as a driving factor.

Figure 11: Grade distributions in the main sample, last traditional (grey) vs. online (black) semesters

(a) 1st semester (b) 2nd semester



Although it is still not possible to exclude with certainty the possibility that graders have been overall more generous when assessing exams sat during online sessions, the descriptive evidence on marginal grades shown in this section speaks against one of the specific forms in which graders' discretion typically manifests.

6.2.2 Assessment type

The fact that grading standards and formats employed to assess students' course performance affects the grade distribution obtained as a result has been well-known for decades, in higher education literature (Powell 1977). Beyond that, research findings reveal that students' perceptions about assessment significantly influence their approaches to learning and studying (Struyven and Janssens 2005), so that additional effects feeding from the assessment methodology into student outcomes are likely.

In order to shed light on changes in assessment type as a possible driving factor behind the aggregate increase in grade averages, we have encoded and analysed assessment information contained in each subject's course plan. Course plans are official documents published by each faculty at the beginning of the academic year: they state the course's content, as well as the type of assessment that will be employed to assign course grades. When Covid hit, a few weeks into the second semester of the 2019 academic year, the Executive Board of the University of Barcelona issued a directive mandating all lecturers in charge of a course during that semester provide a document explaining any changes in the assessment they had planned as a consequence of the pandemic. These exceptional amendments to the course plans were then published by the Faculty of Law and informed students about the new evaluation procedures that would be faced with in each subject.

We have systematised the qualitative information contained in the course plan amendments as well as in the original course plans of the academic semesters between 2018 S1 and 2020 S1, obtaining variables describing the different assessment methods employed in each course, as well as their weights towards the final course grade. Following existing literature on evaluation methods, there are two dimensions of interest that we particularly focus on: difficulty and concentration of course evaluation.

Although it is hard to establish an uncontroversial difficulty ranking in terms of evaluation methods and exam formats employed in higher education, research shows that students hold strong views about different assessment and evaluation formats; they tend to favour multiple-choice format exams and they perform less-well on essay type questions, although there is no difference, in general, in what the two types of tests measure (Struyven and Janssens 2005, Chan and Kennedy 2002, Birenbaum and Feldman 1998, Walstad and W. E. Becker 1994). Based on this literature, we focus our attention on multiple choice exams as well as on group-based assessment as the evaluation methods of interest. Variations in the intensity of the application of these two methods may be interpreted as proxies of variations in assessment leniency and in the amount of individual evaluation pressure weighing on the student.

Table 3 summarises the frequency of occurrence of multiple choice assessment, averaged across all subjects taken by GAP students, distinguishing between traditional and online assessment sessions. These summary statistics show that, as assessments went online and with respect to traditional times, the use of multiple choice tests increased as an examination format. These dynamics are coherent with the fact that that multiple-choice tests are the most intuitive examination format to be transferred to an online format (e.g. through Moodle platforms), as well as the most efficient format to be graded, especially when large student numbers are involved (Nicol 2007, McCoubrie 2004). Table 3 also shows that the use of teamwork assessment activities decreased when evaluation went online, which is again consistent as a response to the mobility restrictions imposed by the pandemic, which made group-study and group-activities less agile to implement.

In sum, the first finding from our analysis of course plans and classification of assessment methodologies is an observable shift in the formats employed as evaluation moved online, which makes the assessment methodology a potential candidate driver explaining changes in grade results. However, the direction in which individual grade results are expected to be impacted by the assessment changes observed is not straightforward to predict, given that our two proxies for 'leniency of evaluation' move in opposite directions, with more frequent multiple-choice testing and less frequent group-assessment. A more precise statistical relationship between our evaluation method proxies and online grade inflation will be explored in the mediation analysis developed in Section 8.

Table 3: Multiple-choice and teamwork assessment before and during online evaluation sessions

	(1) Standard sessions		(2) Online sessions	
	mean	sd	mean	sd
Multiple-choice assessment	0.16	0.37	0.25	0.44
Teamwork assessment	0.33	0.47	0.25	0.44
N	45		44	

Notes: Average occurrence of multiple choice and teamwork as assessment formats across the 45 subjects taught in the GAP degree at the University of Barcelona. Subjects are coded as *1* in a specific semester-year if the given examination format was used in the evaluation package, and *0* otherwise.

6.2.3 Assessment concentration

Like in most other modern higher education institutions, continuous assessment is the default evaluation approach at the University of Barcelona. The final subject grade is usually the result of partial grades obtained from different activities carried out during the academic quarter and of a final exam; the weights carried by the partial components and by the final exam vary by subject.

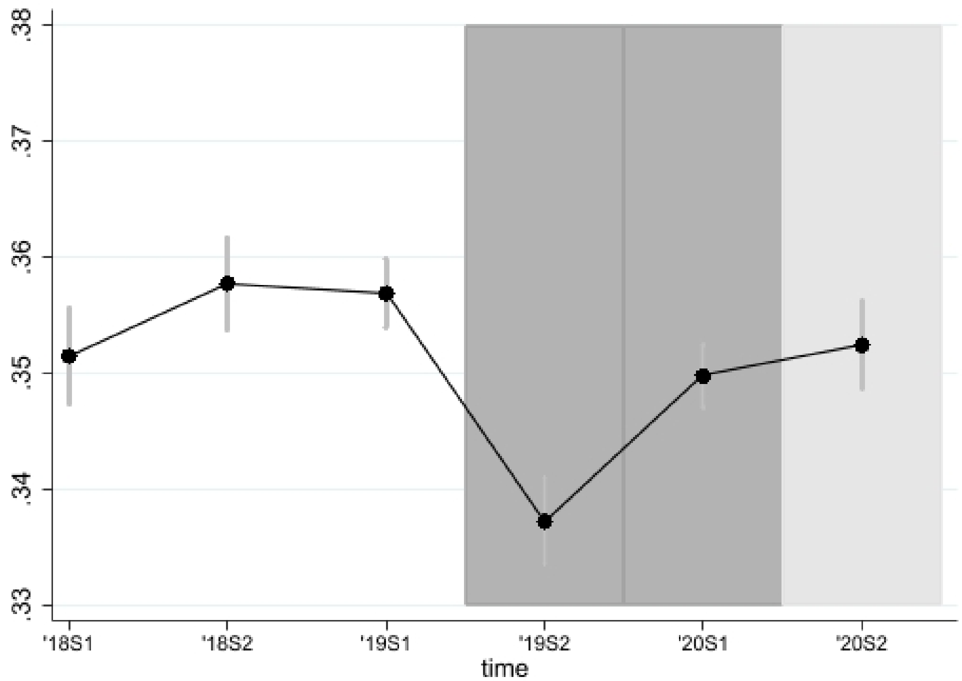
One month after higher education activities had to transition to an online format, the Spanish Commitee of Rectors (CRUE), the Spanish Network of University Quality Agencies (REACU) and the Spanish Ministry of Education issued national guidelines to aid the emergency. Their key recommendations established that evaluation methodologies shall be "considerate of the exceptional learning situation experienced by students"and that a range of different evaluation methods was preferable to a single final exam, explicitly encouraging formative and continuous evaluation approaches (Díez-Gutiérrez and Gajardo Espinoza 2021). In fact, a large national survey ran by Díez-Gutiérrez and Gajardo Espinoza (2021) confirmed Spanish students' preferences for evaluation approaches that employed diverse methods of verification of the acquired knowledge, as recommended by the national guideline documents.

Based on our analysis of course plans and encoding of the evaluation methods used, we calculated the Herfindal-Hirschman index (HHI) characterising assessment concentration in each of the academic semesters between 2018 S1 and 2020 S2. The HHI is a standard measure of concentration, which takes both the number and weight of each component in the set into account. In our setting, this translates into summarising the number of evaluation activities as well as their weight in the continuous assessment scheme.[9] Higher HHI values indicate a more concentrated evaluation system, whereas lower

[9] The range of activities we identified were: class debates, class presentations, case studies, team work assignments, individual assignments, mid-term exams of multiple choice format, mid-term exams with open-ended questions; final exams of multiple choice format; final exams with open-ended questions

values indicate a more dispersed one. Figure 12 plots the HHI values that we obtained for the semesters for which course plans were available.

Figure 12: Average Herfindahl-Hirschman Index (HHI) of subject assessment



The figure shows that semesters in which assessment transitioned online were also characterised by a more dispersed evaluation package, in line with the national recommendations issued for Spain during the pandemic. The drop in concentration was particularly pronounced in the first online semester, 2019 S2, and partially recovered in the following, 2020 S1. It can be noted that the level of concentration in course evaluation had not yet completely recovered to its pre-pandemic level in 2020 S2, a result that is in line with the fact that lecture activities -and possibly a part of continuous evaluation activities- were still held online during that semester.

In conclusion, our quantitative analysis of course plans has identified assessment concentration as a feature that changed during the transition to online evaluation and thus represents a further potential driver of the observed distortions in grade dynamics. Section 8 will shed further light on this hypothesis.

7. Drivers: evidence from survey data

Before turning to the survey data that we collected in June 2020 and exploring drivers behind the observed grade inflation further, we will briefly discuss the suitability of our survey sample to this exercise. In Section 4.2, we commented on the type of selection that we expect in our survey respondents, based on our descriptive statistics (Table A.1): respondents are students who are still actively working on their degree in the academic year 2019/20 and they generally have a higher ability or motivation with respect to the average student from the main sample.

Figure B.1 and Figure B.2 in the Appendix compare the evolution of mean grades and grade standard deviations in the main sample, in the 2019/2020 sample of 'potential respondents' and in the survey sample, over time. As we can see, the most important patterns that were discussed in the previous sections are reproduced both in the 2019/20 sample and in the survey sample: online exam sessions see an increase in average grades with respect to previous years and these fall back again during the first in-classroom session, 2020 S2. In general, results in the survey sample are noisier than in the main sample, due to the smaller number of observations; we attribute the out-of-pattern grade increase in 2018 S1 and 2018 S2 to this cause. A similar pattern can be observed for grade standard deviations, except deviations due to noise: grade dispersion falls during online session and grew back again in 2020 S2. Table A.2 and Figure B.4 in the Appendix reproduce the quantification of the impact of online exam sessions on our survey sample. For 2019 S2 we estimate an approximate 0.2 standard deviation increase in grades with respect to the regular preceding semesters; this magnitude is about half of what was estimated for the main sample but is still strongly significant. For the second online semester, 2020 S1, the magnitude of the grade boost is very similar to the one obtained from the main sample estimation, approximately 0.15 to 0.20 standard deviations compared to the preceding semesters.

In sum, our sample of survey respondents reacted to the online transition in a way closely resembling what is observed for the general student population in our study, save for some overall positive selection in their results and somewhat more noisy estimations of effect magnitudes. This conclusion backs the strategy to leverage our survey sample to draw further insights into channels driving grade dynamics during online evaluation semesters, as shown in the following subsections.

7.1 Teaching and learning quality, student preparation and motivation

Possible causes of grade improvement include more effective teaching and a more favourable learning environment for students. Advantages of online courses in terms of accessibility of learning material and more flexible time management have been documented in several studies previously and prior to the recent pandemic (see systematic reviews in Means et al. 2010 and Coleman and Berge 2018). Our survey explored these dimensions by including questions on students' perception about any changes in teaching and learning quality as these were transferred online.

Figure 13 illustrates the results obtained from classifying students' responses into three groups, depending on whether their perceptions improved, worsened or stayed the same after the transition from in-person to online learning. The questions proposed to students elicited their assessment of the reaction of university faculty to the online transition, a prediction on how distance learning would impact their overall academic results and a prediction on how online continuous evaluation would impact their partial results. It is immediately evident from the plots that the vast majority of respondents did not perceive the online transition as an improvement to their learning experience at the University of Barcelona. In any of the aspects explored, more than 80% of the survey participants perceived a worsening of conditions and result expectations in the new versus the traditional learning environment.

Very similar conclusions to the ones discussed previously can be drawn from survey responses regarding the aspects of academic motivation and perceived preparation. Figure 14 shows that only a small minority of students perceived improvements in terms of mental health and academic stress in the new online setting, and private predictions of grade impacts coming from these dimensions are consistently negative.

Figure 13: Student perceptions on online teaching and learning quality

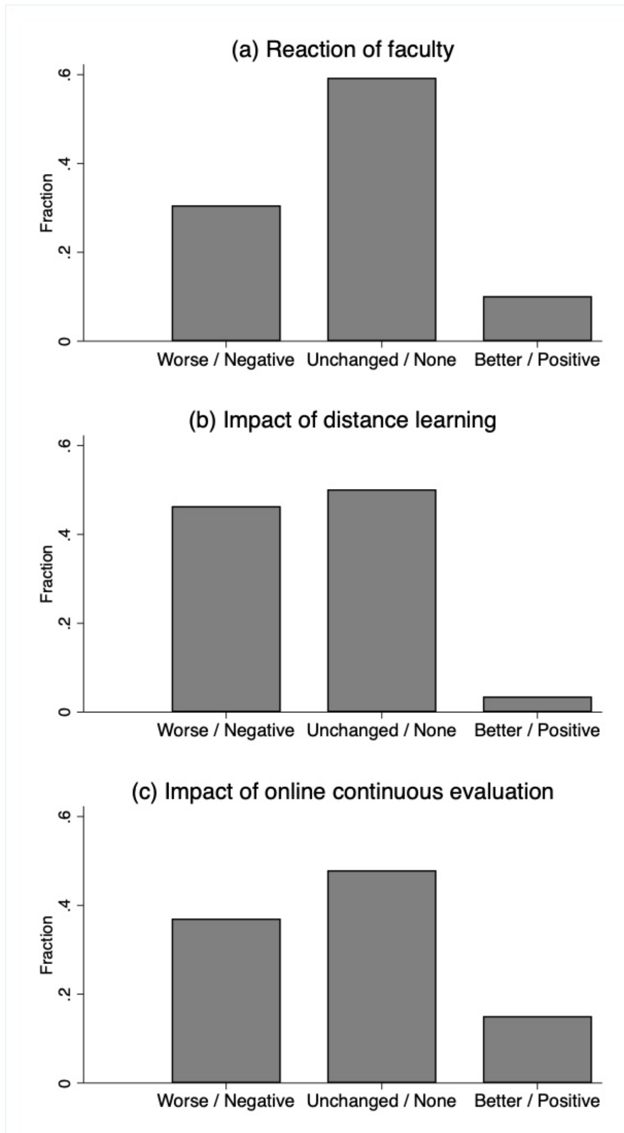
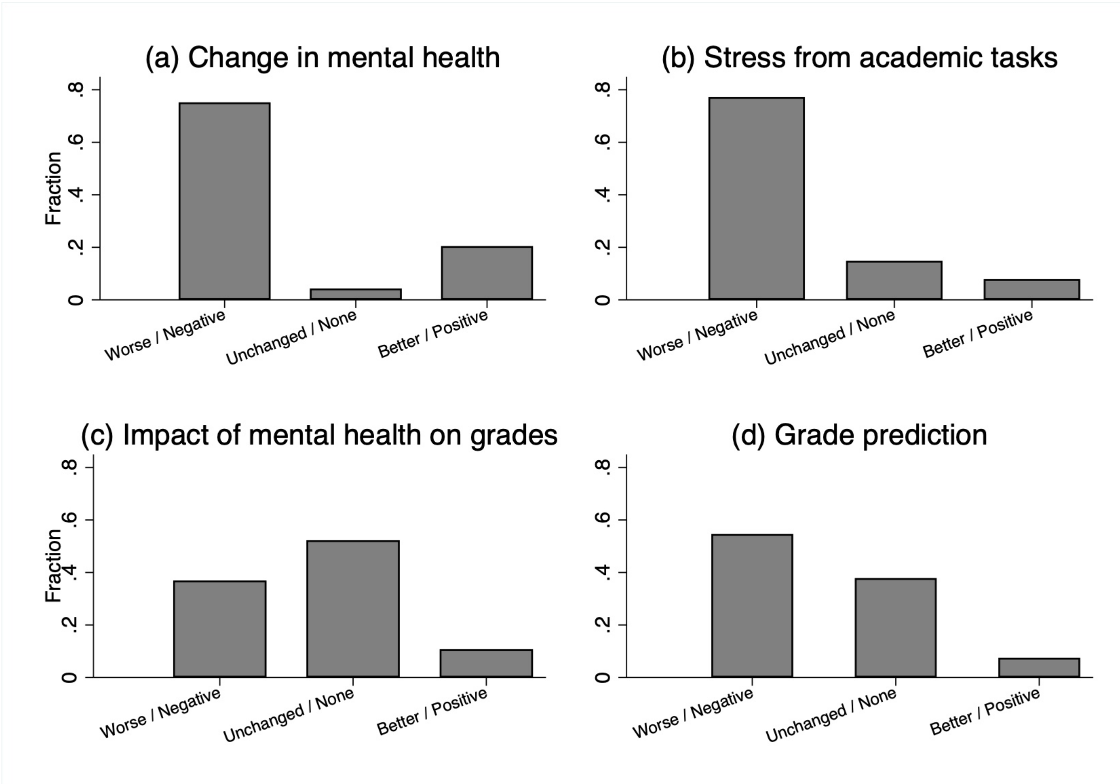


Figure 14: Student perceptions regarding motivation and preparation



On the one hand, this evidence suggests that students perceived the new online setting as more challenging and less comfortable with respect to their standard learning experience at the University, all of which would predict a worsening of academic results during online semesters; on the other hand, though, one could rationalise a strong increase of students' learning effort as a reaction to the new challenges, which may counterbalance the perceived disadvantage and result in a positive effect on grades.

In fact, our survey included question items that allow us to test –and eventually rule out– the latter hypothesis. When asked about the actual time spent engaging with academic activities, 58% of respondents declared an increase in hours of autonomous study during the new online setting, while the remaining 42% reported no change or negative changes in time dedication. Figure B.5 in the Appendix shows that during online sessions there was no significant difference on the evolution of mean grades for these two groups of students, thereby suggesting that the variations in autonomous academic engagement declared were either quantitatively small or were counterbalanced and attenuated by larger variations in the overall learning experience.

7.2 Cheating

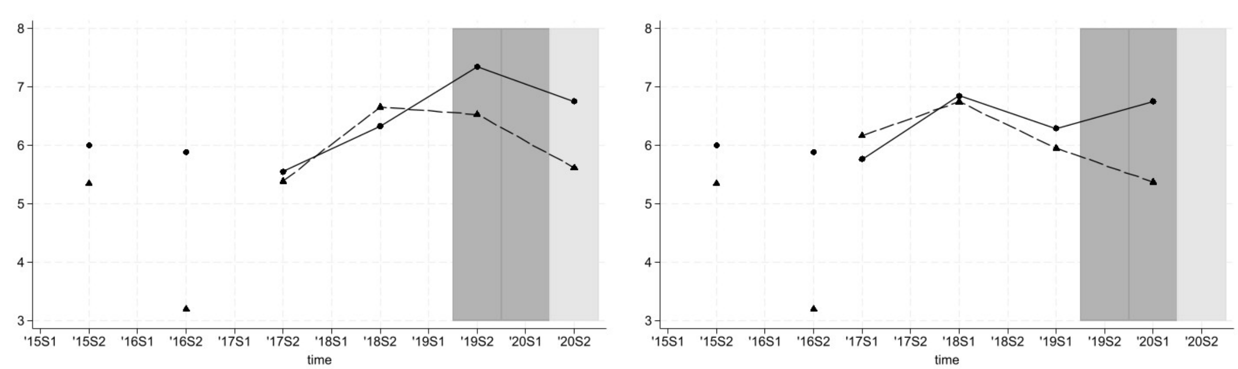
Among the explanations put forward when discussing grade inflation during online exam sessions, there is an increased incidence of cheating behaviour. Online exams often lack human proctoring, so that candidates have the possibility to use notes, the internet, or can even collaborate with other individuals to answer questions; furthermore, detection and punishment are hindered by the fact that online cheating is harder to prove with respect to in-classroom incidents. There are several studies, typically based on self-reports by college students, documenting the fact that cheating is easier under remote examination than in traditional settings (for example, see King and Case 2014; Miller and Young-Jones 2012) and studies detecting widespread cheating during the Spring 2020 online examinations at several universities around the globe (Bilen and Matros 2021; Ebaid 2021). Other papers, however, find inconclusive evidence on the phenomenon (Watson and Sottile 2010; Montejo Bernardo 2020) and conclude that the online testing environment is likely to produce an overall disadvantage to students that offsets greater opportunities to cheat (Fask, Englander and Wang 2014). In sum, it is a priori unclear whether we should expect the cheating dimension to contribute positively to student exam performance or not, in our research setting.

Looking at the survey data collected among GAP students at the University of Barcelona, we are able to explore this issue further. Questions on cheating were included in the survey in an indirect and quite generic manner, in order to minimise discomfort and potential misreporting: students were asked to declare whether they are aware of cheating behaviour that took place during official university exams, without asking about their own behaviour or opinions on the matter. To further limit the scope for concerns about monitoring or similar fears affecting our cheating proxy, we based our analysis on responses concerning the past, regular, pre-pandemic exam sessions.

Based on the answer to the cheating question, we divided surveyed students into a 'cheating-aware' and a 'cheating-unaware' group: the former declared that the occurrence of cheating behaviour was known to them, either sporadically or frequently (56%) while the latter group declared they were unaware of cheating ever taking place during university examinations (44%). When we look at the evolution of mean grades in these two groups separately, as illustrated in Figure 15, an interesting pattern emerges despite the small sample inducing noisiness around results.

Figure 15: Evolution of mean grades among surveyed students, by cheating-awareness (aware solid, unaware dashed).

(a) 2nd semester (b) 1st semester



Note: Mean grades before 2017 by cheating awareness display large amounts of variability.

Students self-declaring as being cheating-aware follow similar grade patterns as their cheating-unaware peers during the academic sessions preceding the pandemic. However, one can appreciate a divergence occurring in the online sessions, with cheating-aware students (solid line in the graph) recording a larger grade increase than their cheating-unaware peers (dashed line in the graph), starting in 2019 S2. In fact, the divergence becomes even more striking in the second online exam session, 2020 S1, in which the average grade recorded for the cheating-unaware group was actually lower than in the previous in-classroom S1 session, so that the whole average grade growth is concentrated in the cheating-aware group.

We interpret this result through the lenses of two implicit assumptions: the first is that, students that declared an awareness about cheating in regular exam sessions are more likely to be involved in cheating behaviour themselves and the second is that they are likely to take better advantage of the greater cheating opportunities offered by the online exam sessions, compared to their 'cheating-unaware' peers. The fact that the second online exam session displays an even stronger gap between cheating-aware and cheating-unaware individuals is in line with this theory, given that cheaters would have by then overcome the surprise and novelties of the new testing environment and would have developed more effective cheating strategies. In the following section, we test cheating among the potential mediators of the grade increases observed during online sessions.

8. Mediation analysis

In this final section of our analysis, we take stock of the evidence collected about potential drivers behind grade inflation in online evaluation and assess their empirical strength through classical mediation analysis. Mediation analysis is widely used in economics to gain insights into the underlying processes that drive the relationships observed in data: it may be used to disentangle total treatment effects into direct and indirect effects or more simply to shed light on the causal mechanisms relating a given variable to an outcome of interest (Chetty et al. 2022; Celli 2022; Huber 2020; Imai, Keele and Tingley 2010). In our setting, we consider online sessions as the 'treatment' of interest and we are interested in determining which of the potential endogenous variables suggested by literature and theory –and for which we have created proxies from either administrative or survey data– are lying in the causal pathway leading to grade inflation.

For this exercise, our baseline specifications of the impact of online semesters on grade averages are those estimated on the basis of the survey sample, originally shown in the second column of Table A.2. These results include individual student controls and thus net out any compositional changes in the population of test takers, changes which were negligible in the first place, according to the findings from Section 6.1. The baseline estimations are reported again in the first column of Table 4 and the following columns show how results were impacted as variables proxying the different potential 'mediators' were alternatively introduced.

Models (2) and (3) in Table 4 show the results of introducing variables encoding individual student perceptions in terms of own preparation and academic motivation, or in terms of online teaching quality. Such perceptions were discussed in Sections 7.1 and were found to have been negatively impacted by the transition to online instruction and examination. We observe that our baseline results are left essentially unchanged by the introduction of proxies for student preparation and motivation (see Model 2), which yields the conclusion that changes in the perceived individual mental health, academic preparation and stress levels do not appear to mediate the relationship between online evaluation and grade inflation observed. Conversely, the grade inflation impact estimated for online sessions is actually magnified as perceived teaching quality proxies are introduced (see Model 3), the interpretation of which is that perceived challenges around the learning environment, such as available resources and teaching quality, actually appear to act as 'suppressors' in the relationship of interest (MacKinnon 2012). That is to say, one possible interpretation of this evidence is that grade boosts during online-evaluation semesters would be even larger, removing the compensating effects of the more challenging teaching and learning environment.

Models (4) and (5) in Table 4 show the results obtained when including the proxies that we have constructed for assessment methods and assessment concentration faced by students in each evaluation session and subject taken.

Table 4: How different variables mediate the relationship between evaluation sessions and grade averages

	(1) Baseline	(2) Student prep. & motivation	(3) Teaching quality	(4) Assessment method	(5) Assessment concentration	(6) Cheating
2016 S1	0.00	0.00	0.00	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
2017 S1	0.02	0.01	0.02	0.01	0.02	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
2018 S1	0.22***	0.22***	0.22***	0.22***	0.22***	0.23***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
2019 S1	0.17***	0.17***	0.17***	0.17***	0.17***	0.17***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
2020 S1	0.29***	0.27***	0.50***	0.38***	-0.10	0.06***
	(0.01)	(0.04)	(0.06)	(0.04)	(0.35)	(0.02)
S2	-0.28***	-0.29***	-0.28***	-0.28***	-0.28***	-0.28***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
2016 S2	0.24***	0.24***	0.24***	0.24***	0.24***	0.24***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
2017 S2	0.26***	0.26***	0.26***	0.26***	0.26***	0.26***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
2018 S2	0.21***	0.22***	0.22***	0.22***	0.21***	0.21***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
2019 S2	0.45***	0.44***	0.68***	0.53***	0.10	0.24***
	(0.01)	(0.04)	(0.06)	(0.06)	(0.31)	(0.02)
2020 S2	0.19***	0.20***	0.19***	0.25***	0.19***	0.21***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Female	0.08	0.06	0.06	0.08	0.08	0.11
	(0.06)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)
Spanish	0.34***	0.32***	0.36***	0.34***	0.35***	0.36***
	(0.06)	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)
Access grade	0.25***	0.26***	0.25***	0.25***	0.25***	0.25***
	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)
Student prep. & motivation	No	Yes	No	No	No	No
Teaching quality	No	No	Yes	No	No	No
Assessment method	No	No	No	Yes	No	No
Assessment concentration	No	No	No	No	Yes	No
Cheating	No	No	No	No	No	Yes
Mean(y)	0.25	0.25	0.25	0.25	0.25	0.25
sd(y)	0.89	0.89	0.89	0.89	0.89	0.89
N.obs	2,065	2,058	2,065	2,065	2,065	2,065
R-squared	0.10	0.12	0.12	0.12	0.10	0.12

Notes: Linear regression results of exam grades on academic semester fixed effects and a set of student observables. The dependent variable is individual exam grades expressed in pre-Covid, full-sample exam-level standard deviations. SEs clustered at the year-semester level in parentheses. p < 0.10 *, < 0.05 **, < 0.01 ***.

In Section 6.2.2, we discussed assessment formats as a potential driver of results, and a particularly hard one to evaluate, due to the complex nature of assessment practices and the disagreement around what is 'easier' or 'more difficult' for students. The evidence that we were able to collect on the less-controversial proxies for evaluation leniency, the use of multiple-choice and group-work assessment, moved in opposing directions, calling for a more formal exploration of the statistical relationships between these variables and grade averages. The results in column (4) suggest that the changes that we observe in the subject evaluation packages, and in the use of formats that are traditionally considered 'less demanding' for the individual more specifically, do not explain and actually appear to partially suppress the relationship between online evaluation and grade inflation.

Our first positive result in terms of mediating factors that explain the relationship between online evaluation and grade averages is obtained when we introduce assessment concentration indicators to our regression model. In Section 6.2.3, we showed that faculty of the GAP degree at the University of Barcelona appears to have followed the national recommendations mandating the application of diverse and dispersed evaluation practices during the pandemic, avoiding one-off examinations carrying a high weight in terms of final course result as much as possible. Column (5) of Table 4 shows how the estimation of the 'online evaluation' effect, captured by the coefficients on the academic semester dummies 2019 S2 and 2020 S1, drops dramatically in magnitude and loses statistical significance when our dispersion measures, subject-semester level HHI indices, are added to the baseline specification. This can be interpreted as evidence of the higher dispersion in assessment activities recorded during online evaluation semesters being a factor mediating the impact in terms of the grade average boosts observed, a result in line not only with the motivation behind the national guidelines issued by Spanish education authorities during the Covid pandemic, but also with a large body of education and cognitive science literature highlighting the learning benefits deriving from continuous assessment and 'distributed learning' practices (see, for example Wiseheart et al. 2019, Budé et al. 2011, Benjamin and Tullis 2010).

The last column of Table 4 augments our baseline model a variable coding the individual responses about 'cheating awareness'. In Section 7.2, we discussed cheating as a common problem affecting remote examination practices in particular, and we showed descriptive evidence of a grade-gap opening between students that assessed themselves as being 'cheating aware,' who appear to have seen the largest grade boosts during online sessions, and those who declared not to know of any cheating happening during university examinations. When we introduce individual responses about cheating to our regression model of grade averages, both the 2019 S2 and the 2020 S1 coefficients drop in magnitude, which can be interpreted as evidence for a mediation effect running through our cheating proxy.

The mediation analysis developed in this section complemented the descriptive evidence from the previous sections by providing empirical guidance about which candidate mechanisms to either endorse or discard as likely drivers of the grade inflation observed during online examination sessions. We found student preparation and academic motivation, the quality of the teaching and learning environment and the assessment formats used are unlikely to explain observed grade increases in connection to online assessment; on the contrary, we find that, if anything, some of these factors push towards a reduction in the grade boost. Conversely, we found that changes in assessment concentration and the extent of cheating during evaluations provide a statistical explanation for why grades are higher in the two semesters in which learning and evaluation practices were administered online.

9. Discussion and concluding remarks

In this paper, we have analysed the reaction of university grade distributions when course learning and evaluation activities transition to an online format. The main descriptive result is a substantial grade inflation and a reduction in dispersion, mostly driven by an improvement in the results of students belonging to the lower part of the performance distribution. In terms of heterogeneity, we also find evidence of a comparative disadvantage for students from lower socioeconomic background but discard any considerable gender differences. Despite our analysis not relying on an exceptionally large sample size, these descriptive findings align with and corroborate those of existing studies.

An important supplementary descriptive deduction that we are able to make, thanks to the features of our specific setting, is that grade boosts and lower dispersion are attributable to online *examination* in specific, and not to online *instruction* in general. Our last academic semester of observation at the University of Barcelona still featured online instruction in combination with traditional, face-to-face evaluation: in this semester, we observed grade averages and any other distributional features, such as grade dispersion and achievement gaps between student types, reverting back to traditional levels.

While this is not the first study documenting grade inflation in higher education during the recent pandemic, we have contributed to the existing literature by providing a systematic scrutiny of potential drivers behind the grade dynamics observed. This was made possible by leveraging administrative data, survey data and data from course plans collected at the University of Barcelona, a highly valuable combination of sources not usually available to researchers.

We are able to discard compositional effects in the pool of exam candidates as drivers of the grade inflation observed and we do not find empirical support for targeted 'generosity' during the grading process, although we are not able to exclude the hypothesis of overall 'easier' exam content having been administered or more sophisticated forms of leniency by graders.

The systematisation of course plans that we have carried out allowed us to also explore assessment formats and the degree of assessment dispersion during term time as potential drivers behind inflation. We find evidence for changes in both of these dimensions during online sessions, but only the increase in assessment dispersion provides a statistical explanation for the higher grades observed. The second driver that we find empirical support for this grade increase is cheating behaviour during online examinations.

Our reading of the overall picture delivered by this analysis is that transferring evaluation activities to an online format has induced grade improvements, albeit mostly for traditionally lower-performing students –with a relative disadvantage for those from lower socioeconomic backgrounds and worse technology at home. Our data identifies two features that came along with the online transition as channels that help explain the higher grade results: the broader spectrum of assessment opportunities during the academic semester, but also the larger room for cheating during remote evaluation.

In terms of policy advice, our findings reinforce the existing support for continuous assessment as a preferable method to be employed in higher education courses, as weaker students in particular appear to benefit from the option to demonstrate their understanding and skills through various channels, thereby catering to different learning styles and abilities. The second policy implication, as it was widely implemented during Covid times, is to provide students with adequate equipment for on-line learning activities. Finally, a straightforward urgency emphasised by our analysis is the implementation of robust academic integrity measures to address the increased cheating opportunities offered by online evaluations. This can include the use of plagiarism detection software, proctored exams, and promoting a culture of academic honesty through educational initiatives and clear communication of expectations.

Both our descriptive analysis and the inspection of mechanisms through the mediation framework delivered the observation that student performance appears to have suffered due to the lower perceived quality of the online learning, teaching and evaluation environments with respect to traditional delivery methods. This calls for targeted support and resources for academic skills development, particularly in remote settings, as well as for providing faculty members with training and support in designing effective online teaching and meaningful assessments that accurately measure student learning outcomes.

Finally, while the transition to online evaluation has shown positive outcomes for lower-performing students, we also documented a comparative loss in results for individuals that are likely to belong to more disadvantaged socioeconomic backgrounds, as proxied by their scholarship status. Consequently, the need to ensure equitable access to online resources and to implement measures bridging any digital divide that may contribute to the observed comparative disadvantaged during remote delivery of academic activities is worth stressing.

Beyond specific conclusions, our analysis has demonstrated the value of having access to granular data from diverse sources which allow for precisely quantified aggregate dynamics to be obtained, but also data-driven insights on the mechanisms driving the same. A comprehensive set of results of this kind represents a high-quality foundation for the refinement of policies and practices in higher education, including the case of transitions to online versions thereof, striving to ensure an accurate measurement of individual performance and the equitable treatment of all students.

References

Andrew, Alison et al. (2020). "Inequalities in Children's Experiences of Home Learning during the COVID-19 Lockdown in England". In: Fiscal Studies 41.3.

Astorne-Figari, Carmen and Jamin D Speer (2019). "Are changes of major major changes? The roles of grades, gender, and preferences in college major switching". In: Economics of Education Review 70, pp. 75–93.

Aucejo, Esteban M et al. (Nov. 2020). "The impact of COVID-19 on student experiences and expectations: Evidence from a survey". In: Journal of Public Economics 191.

Ayllón, Sara (Aug. 2022). "Online teaching and gender bias". In: Economics of Education Review 89.

Azubuike, Obiageri Bridget, Oyindamola Adegboye and Habeeb Quadri (2021). "Who gets to learn in a pandemic? Exploring the digital divide in remote learning during the COVID-19 pandemic in Nigeria". In: International Journal of Educational Research Open 2, p. 100022.

Barnum, Matt and Carmen Bryan (June 2020). "America's great remote-learning experiment: What surveys of teachers and parents tell us about how it went". In: Chalkbeat.

Benjamin, Aaron S. and Jonathan Tullis (2010). "What makes distributed practice effective?" In: Cognitive Psychology 61.3, pp. 228–247. ISSN: 0010-0285. DOI: https://doi.org/10.1016/j.cogpsych.2010.05.004. URL: https://www.sciencedirect.com/science/article/pii/S0010028510000332.

Bilen, Emre and Alexander Matros (2021). "Online cheating amid Covid-19". In: Journal of Economic Behavior & Organization 182, pp. 196–211.

Bird, Kelsey A, Benjamin L Castleman and Gina Lohner (Apr. 2022). "Negative Impacts From the Shift to Online Learning During the COVID-19 Crisis: Evidence From a Statewide Community College System". In: AERA Open.

Birenbaum, Menucha and Rose A. Feldman (1998). "Relationships between learning patterns and attitudes towards two assessment formats". In: Educational Research 1.40, pp. 90–98.

Bowen, William G, Matthew M Chingos and Katharine A Lac (2014). "Interactive Learning Online at Public Universities: Evidence from a Six-Campus Randomized Trial". In: Journal of Policy Analysis and Management 33.1, pp. 94–111.

Browning, M H et al. (Jan. 2021). "Psychological impacts from COVID-19 among university students: Risk factors across seven states in the United States". In: PLoS One 16.1.

Budé, Luc et al. (2011). "The effect of distributed practice on students'conceptual understanding of statistics". In: Higher Education 62.1, pp. 69–79. DOI: 10.1007/s10734-010-9366-y. URL: https://doi.org/10.1007/s10734-010-9366-y.

Celli, Viviana (2022). "Causal mediation analysis in economics: Objectives, assumptions, models". In: Journal of Economic Surveys 36.1, pp. 214–234.

Chan, Nixon and Peter E Kennedy (2002). "Are multiple-choice exams easier for economics students? A comparison of multiple-choice and "equivalent" constructed-response exam questions". In: Southern Economic Journal 68.4, pp. 957–971.

Chetty, Raj et al. (2022). "Social capital I: measurement and associations with economic mobility". In: Nature 608.7921, pp. 108–121.

Coleman, Mariann and Zane L Berge (2018). "A review of accessibility in online higher education". In: Online Journal of Distance Learning Administration 21.1, pp. 1–7.

Correia, Ana-Paula (2020). "Healing the digital divide during the COVID-19 pandemic." In: Quarterly Review of Distance Education 21.1.

Devkota, Kamal Raj (Mar. 2021). "Inequalities reinforced through online and distance education in the age of COVID-19: The case of higher education in Nepal". In: International Review of Education 67, pp. 145–165.

Díez-Gutiérrez, Enrique and Katherine Gajardo Espinoza (2021). “Online assessment in higher education during Spanish confinement by COVID-19: The view of students”. In: Journal of University Teaching Learning Practice 5.18, pp. 709–712.
Ebaid, Ibrahim El-Sayed (2021). “Cheating among Accounting Students in Online Exams during Covid-19 Pandemic: Exploratory Evidence from Saudi Arabia”. In: Asian Journal of Economics, Finance and Management 4.1, pp. 9–19.
El Correo Web (2022). La próxima Selectividad seguirá siendo igual de fácil. https://elcorreoweb.es/andalucia/la-proxima-selectividad-seguira-siendo-igual-de-facil-NG7585539. Accessed: [Insert Date].
El Español (2022). “Fáciles, los exámenes de la EBAU este año: así es la Selectividad que no hace sudar a los estudiantes”. In: El Español. URL: https://www.elespanol.com/reportajes/20220611/faciles-examenes-ebau-ano-selectividad-hace/679182424_0.html.
El País (2022). La escalada vertiginosa de notas en Bachillerato. https://elpais.com/educacion/secundaria-bachillerato-fp/2022-07-10/la-escalada-vertiginosa-de-notas-en-bachillerato-los-sobresalientes-de-los-que-llegan-a-selectividad-se-doblan-en-seis-anos.html. Accessed: [July 2023].
Eruchalu, Chukwuma N et al. (2021). “The expanding digital divide: digital health access inequities during the COVID-19 pandemic in New York City”. In: Journal of Urban Health 98, pp. 183–186.
Europa Press (2021). “Las dos próximas EBAU continuarán en formato pandemia con más opciones y respuesta a preguntas tipo test”. In: Europa Press. URL: https://www.europapress.es/sociedad/noticia-dos-proximas-ebau-continuaran-formato-pandemia-mas-opciones-respuesta-preguntas-tipo-test-20211210075955.html.
Fask, Andrew, Freda Englander and Zixuan Wang (2014). “Do Online Exams Facilitate Cheating? An Experiment Designed to Separate Possible Cheating from the Effect of the Online Test Taking Environment”. In: Journal of Academic Ethics 12, pp. 101–112.
Freixa Niella, Montserrat, Juan Llanes Ordóñez and Miquel Venceslao Pueyo (2018). “Abandonment in the Student’s educational itinerary of ADE of the University of Barcelona”. In: Revista de Investigación Educativa 36.1, pp. 185–202.
Hamm, Jeremy M et al. (2019). “A motivation treatment to enhance goal engagement in online learning environments: Assisting failure-prone college students with low optimism.” In: Motivation Science 5.2, p. 116.
Hansen, Anne Toft, Ulrik Hvidman and Hans Henrik Sievertsen (2024). “Grades and employer learning”. In: [Forthcoming] Journal of Labor Economics.
Huber, Martin (2020). “Mediation analysis”. In: Handbook of labor, human resources and population economics, pp. 1–38.
Imai, Kosuke, Luke Keele and Dustin Tingley (2010). “A general approach to causal mediation analysis”. In: Psychological Methods 4.15, pp. 309–334.
Jaeger, David A et al. (2021). “The global covid-19 student survey: First wave results”. In:
Karadag, Engin (2021). “Effect of COVID-19 pandemic on grade inflation in higher education in Turkey”. In: PLoS ONE 16.8. DOI: https://doi.org/10.1371/journal.pone.0256688.
King, David L and Christine J Case (2014). “E-Cheating: Incidence and Trends Among College Students”. In: Issues in Information Systems 15.1, pp. 20–27.
Kofoed, Michael et al. (2021). “Zooming to class?: Experimental evidence on college students&Apos; online learning during Covid-19”. In: Online Learning During COVID-19. IZA Discussion Paper 14356.
MacKinnon, David (2012). Introduction to statistical mediation analysis. Routledge.
McCoubrie, Paul (2004). “Improving the fairness of multiple-choice questions: a literature review”. In: Medical Teacher 8.26, pp. 709–712.
Means, Barbara et al. (2010). Evaluation of evidence-based practices in online learning. Tech. rep. Office of Planning, Evaluation, Policy Development, Policy and Program Studies Service, US Department of Education.

Miller, Alison and Amanda D Young-Jones (2012). "Academic integrity: Online classes compared to face-to-face classes". In: Journal of Instructional Psychology 39.3, pp. 138–145.
Montejo Bernardo, J. (2020). "Exámenes no presenciales en época del COVID-19 y el temor al engaño. Un estudio de caso en la Universidad de Oviedo". In: Magister 31.1, pp. 102–110.
Nicol, David (2007). "E-assessment by design: using multiple-choice tests to good effect". In: Journal of Further and Higher Education 1.31, pp. 53–64.
OECD (2021). Education Policy Outlook 2021, p. 251. DOI: https://doi.org/https://doi.org/10.1787/75e40a16-en. URL: https://www.oecd-ilibrary.org/content/publication/75e40a16-en.
Orlov, George et al. (2021). "Learning during the COVID-19 pandemic: It is not who you teach, but how you teach". In: Economics Letters 202, p. 109812.
Patrinos, Harry Anthony, Emiliana Vegas and Rachel Carter-Rau (2022). "An analysis of COVID-19 student learning loss". In:
Piopiunik, Marc et al. (2020). "Skills, signals, and employability: An experimental investigation". In: European Economic Review 123, p. 103374.
Powell, Robert W. (1977). "Grades, learning, and student evaluation of instruction". In: Research in Higher Education 7.3, pp. 193–205. DOI: 10.1007/BF00991986. URL: https://doi.org/10.1007/BF00991986.
Ramsetty, Anita and Cristin Adams (2020). "Impact of the digital divide in the age of COVID-19". In: Journal of the American Medical Informatics Association 27.7, pp. 1147–1148.
Rask, Kevin (2010). "Attrition in STEM fields at a liberal arts college: The importance of grades and pre-collegiate preferences". In: Economics of education review 29.6, pp. 892–900.
Respondek, Lisa et al. (2020). "Linking changes in perceived academic control to university dropout and university grades: A longitudinal approach." In: Journal of Educational Psychology 112.5, p. 987.
Robbins, Steven B et al. (2009). "Intervention effects on college performance and retention as mediated by motivational, emotional, and social control factors: integrated meta-analytic path analyses." In: Journal of Applied Psychology 94.5, p. 1163.
Rodríguez-Planas, Núria (2022a). "COVID-19, college academic performance, and the flexible grading policy: A longitudinal analysis". In: Journal of Public Economics 207.
— (Apr. 2022b). "Hitting where it hurts most: COVID-19 and low-income urban college students". In: Economics of Education Review 87.
Rosendale, J. A. (May 2017). "Gauging the value of MOOCs: An examination of American employers' perceptions toward higher education change". In: Higher Education, Skills and Work-Based Learning 7.2.
Struyven Katrien, Filip Dochy and Steven Janssens (2005). "Students' perceptions about evaluation and assessment in higher education: a review1". In: Assessment & Evaluation in Higher Education 30.4, pp. 325–341. DOI: 10.1080/02602930500099102.
Tillinghast, Jonathan A., Dr. James W. Mjelde and Anna Yeritsyan (2023). "COVID-19 and Grade Inflation: Analysis of Undergraduate GPAs During the Pandemic". In: SAGE Open 13.4, p. 21582440231209110. DOI: 10.1177/21582440231209110.
Toro, D. (July 2022). "Academic performance, grade inflation and regulatory exceptions in higher education during the pandemic". In: ICERI2022 Proceedings. 15th annual International Conference of Education, Research and Innovation. Seville, Spain: IATED, pp. 3990–3994. ISBN: 978-84-09-45476-1. DOI: 10.21125/iceri.2022.0969. URL: https://doi.org/10.21125/iceri.2022.0969.
Vallespín, Inma (May 2020). "Las universidades detectan 615 alumnos que no pueden seguir el curso 'online'". In: El País.
Walstad, William B and William E Becker (1994). "Achievement differences on multiple-choice and essay tests in economics". In: The American Economic Review 84.2, pp. 193–196.
Watson, Gregory R. and James Sottile (2010). "Cheating in the Digital Age: Do Students Cheat More in Online Courses?" In: Online Journal of Distance Learning Administration 13.1.

Whiting, Kate (2020). World Economic Forum - Is this what higher education will look like in 5 years? IPSOS Surveys, https://www.weforum.org/agenda/2020/11/higher-education-online-change-cost-covid-19/ [Accessed: (29/01/2024)].

Wiseheart, Melody et al. (2019). "The Cambridge handbook of cognition and education". In: Cambridge University Press. Chap. Enhancing the Quality of Student Learning Using Distributed Practice.

Xu, Di and Shanna Smith Jaggars (2013). "The impact of online learning on students' course outcomes: Evidence from a large community and technical college system". In: Economics of Education Review 37, pp. 46–57.

Appendix A. Tables

Table A.1: Student-level descriptives

	(1) Main sample		(2) Students in 2019		(3) Survey sample	
	mean	sd	mean	sd	mean	sd
Gender	0.55	0.50	0.55	0.50	0.69	0.47
Age	21.47	6.41	20.92	5.62	21.65	6.63
Individual GPA	5.85	1.72	6.16	1.31	6.30	1.31
Perc. exam absences	2.37	9.64	0.96	5.87	0.43	4.15
University access grade	6.20	0.92	6.23	0.88	6.31	0.87
Scholarship status	0.35	0.48	0.38	0.49	0.41	0.49
N	766		331		93	

Table A.2: Impact of online exam sessions on grade averages in the survey sample

	(1) Baseline	(2) Individual controls	(3) Subject FEs
2016 S1	-0.04***	0.00	-0.02
	(0.00)	(0.01)	(0.02)
2017 S1	-0.04***	0.02	-0.06**
	(0.00)	(0.01)	(0.02)
2018 S1	0.20***	0.22***	0.20***
	(0.00)	(0.01)	(0.03)
2019 S1	0.17***	0.29***	0.30***
	(0.00)	(0.01)	(0.06)
2020 S1	0.29***	0.29***	0.30***
	(0.00)	(0.01)	(0.06)
S2	-0.26***	-0.28***	-0.05
	(0.00)	(0.01)	(0.33)
2016 S2	0.20***	0.24***	0.26***
	(0.00)	(0.01)	(0.03)
2017 S2	0.22***	0.26***	0.25***
	(0.00)	(0.01)	(0.03)
2018 S2	0.17***	0.45***	0.48***
	(0.00)	(0.01)	(0.05)
2019 S2	0.40***	0.45***	0.48***
	(0.00)	(0.01)	(0.05)
2020 S2	0.18***	0.19***	0.03
	(0.00)	(0.01)	(0.09)
Age		0.00	-0.00
		(0.00)	(0.00)
Female		0.08	0.07
		(0.06)	(0.05)
Spanish		0.34***	0.31***
		(0.06)	(0.06)
Access grade		0.25***	0.25***
		(0.03)	(0.04)
Subject FEs	No	No	Yes
Mean(y)	0.25	0.25	0.25
sd(y)	0.89	0.89	0.89
N.obs	2,071	2,065	2,065
R-squared	0.03	0.10	0.20

Notes: Linear regression results of exam grades on academic semester fixed effects and a set of student observables and subject fixed effects. The dependent variable is individual exam grades expressed in pre-Covid, full-sample exam-level standard deviations. SEs clustered at the year-semester level in parentheses. p < 0.10 *, < 0.05 **, < 0.01 ***.

Appendix B. Figures

Figure B.1: Evolution of mean grades. Main sample (solid), surveyed students (dashed) and students enrolled in 2019 (dash-dotted)

(a) 2nd semester (b) 1st semester

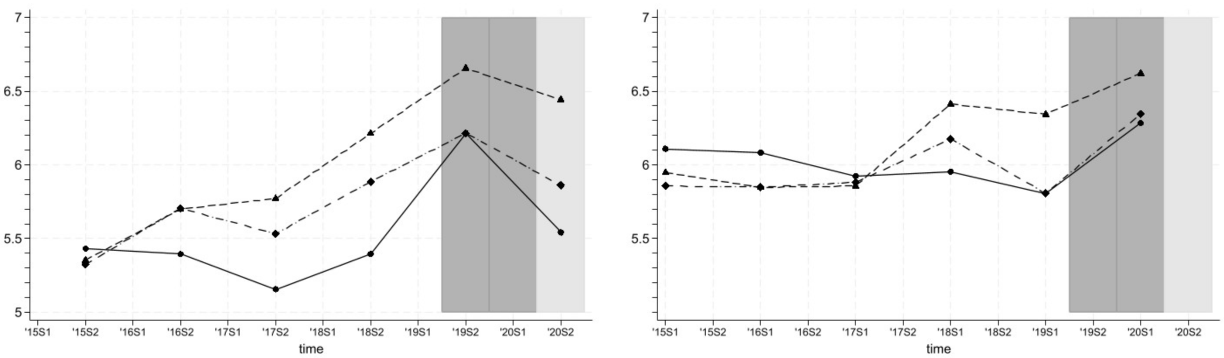


Figure B.2: Evolution of grade standard deviations. Main sample (solid), surveyed students (dashed) and students enrolled in 2019 (dash-dotted)

(a) 2nd semester (b) 1st semester

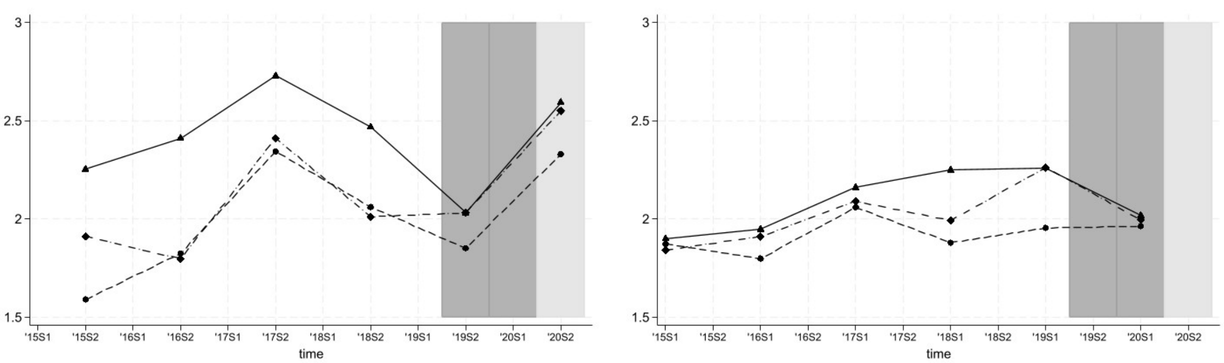
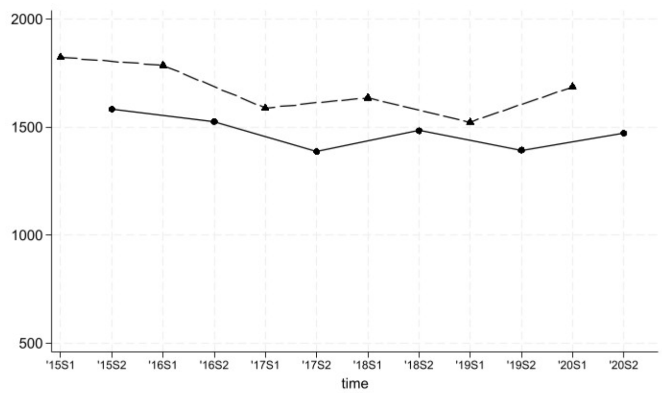
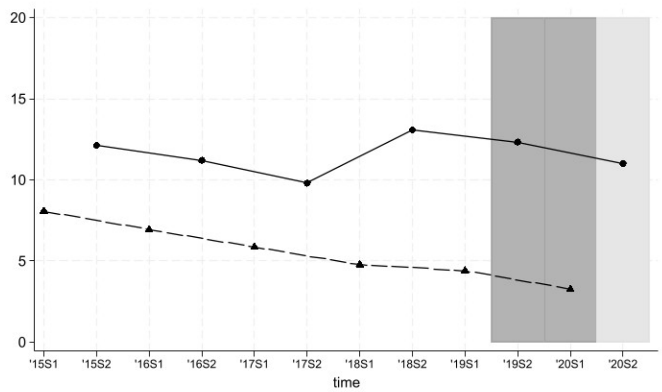


Figure B.3: Evolution of number of test takers, share of no-shows and number of credits obtained in the main sample. First semester dashed, second semester solid.

(a) Test takers



(b) Exam no-shows



(c) Credits obtained

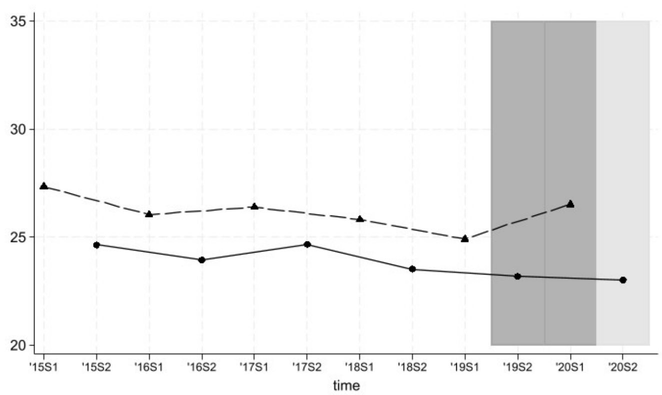


Figure B.4: Estimation of academic semester fixed effects in the survey sample, holding constant student observables

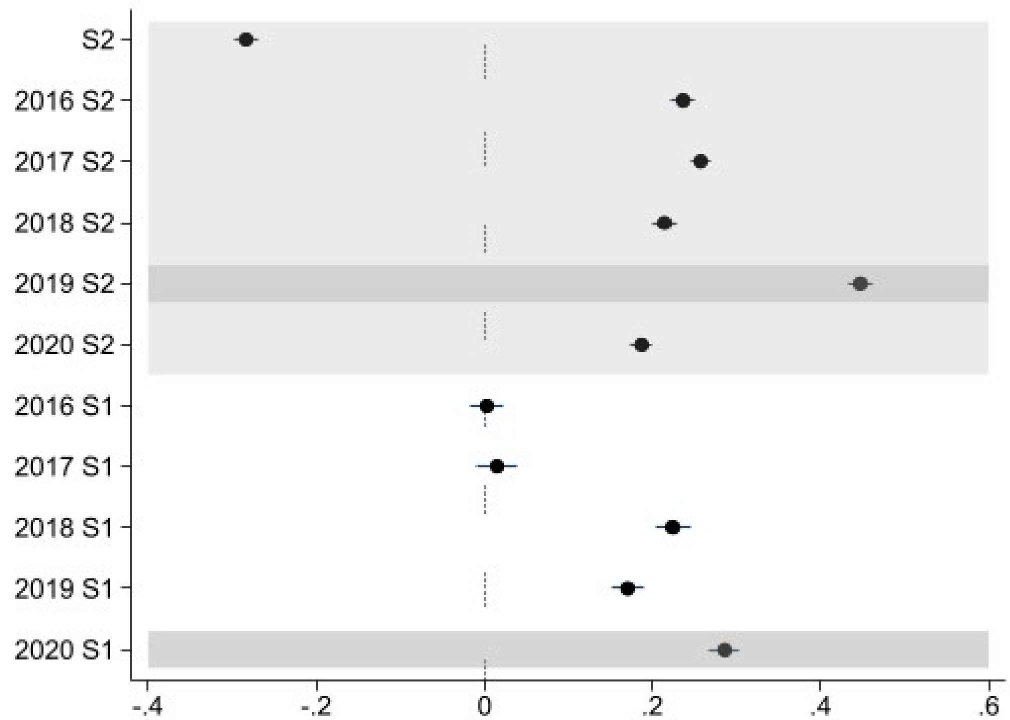
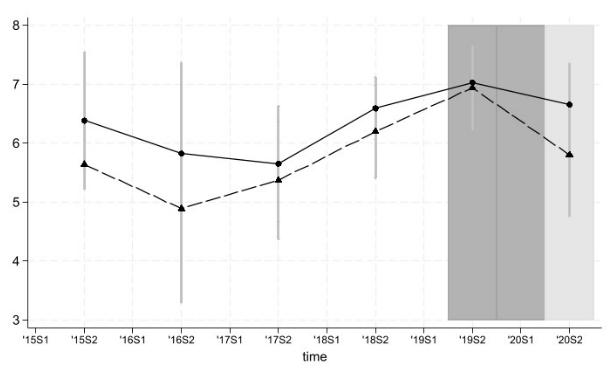
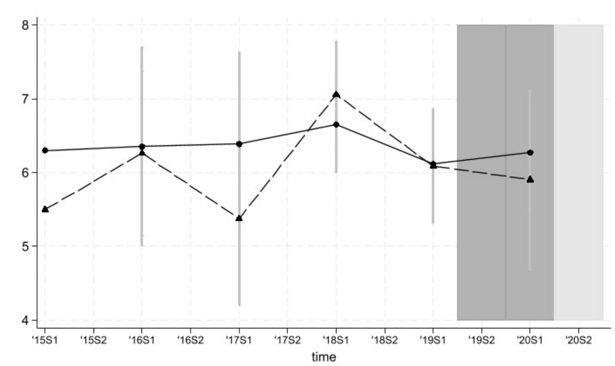


Figure B.5: Evolution of mean grades among surveyed students, by change in the hours devoted to studying in online times. Increased solid, equal or reduced dashed.

(a) 2nd semester



(b) 1st semester



2019

2019/1, Mediavilla, M.; Mancebón, M. J.; Gómez-Sancho, J. M.; Pires Jiménez, L.: "Bilingual education and school choice: a case study of public secondary schools in the Spanish region of Madrid"
2019/2, Brutti, Z.; Montolio, D.: "Preventing criminal minds: early education access and adult offending behavior"
2019/3, Montalvo, J. G.; Piolatto, A.; Raya, J.: "Transaction-tax evasion in the housing market"
2019/4, Durán-Cabré, J.M.; Esteller-Moré, A.; Mas-Montserrat, M.: "Behavioural responses to the reintroduction of wealth taxes. Evidence from Spain"
2019/5, Garcia-López, M.A.; Jofre-Monseny, J.; Martínez Mazza, R.; Segú, M.: "Do short-term rental platforms affect housing markets? Evidence from Airbnb in Barcelona"
2019/6, Domínguez, M.; Montolio, D.: "Bolstering community ties as a means of reducing crime"
2019/7, García-Quevedo, J.; Massa-Camps, X.: "Why firms invest (or not) in energy efficiency? A review of the econometric evidence"
2019/8, Gómez-Fernández, N.; Mediavilla, M.: "What are the factors that influence the use of ICT in the classroom by teachers? Evidence from a census survey in Madrid"
2019/9, Arribas-Bel, D.; Garcia-López, M.A.; Viladecans-Marsal, E.: "The long-run redistributive power of the net wealth tax"
2019/10, Arribas-Bel, D.; Garcia-López, M.A.; Viladecans-Marsal, E.: "Buildings and cities: delineating urban areas with a machine learning algorithm"
2019/11, Bordignon, M.; Gamalerio, M.; Slerca, E.; Turati, G.: "Stop invasion! The electoral tipping point in anti-immigrant voting"

2020

2020/01, Daniele, G.; Piolatto, A.; Sas, W.: "Does the winner take it all? Redistributive policies and political extremism"
2020/02, Sanz, C.; Solé-Ollé, A.; Sorribas-Navarro, P.: "Betrayed by the elites: how corruption amplifies the political effects of recessions"
2020/03, Farré, L.; Jofre-Monseny; J., Torrecillas, J.: "Commuting time and the gender gap in labor market participation"
2020/04, Romarri, A.: "Does the internet change attitudes towards immigrants? Evidence from Spain"
2020/05, Magontier, P.: "Does media coverage affect governments'preparation for natural disasters?"
2020/06, McDougal, T.L.; Montolio, D.; Brauer, J.: "Modeling the U.S. firearms market: the effects of civilian stocks, crime, legislation, and armed conflict"
2020/07, Veneri, P.; Comandon, A.; Garcia-López, M.A.; Daams, M.N.: "What do divided cities have in common? An international comparison of income segregation"
2020/08, Piolatto, A.: "'Information doesn't want to be free': informational shocks with anonymous online platforms"
2020/09, Marie, O.; Vall Castello, J.: "If sick-leave becomes more costly, will I go back to work? Could it be too soon?"
2020/10, Montolio, D.; Oliveira, C.: "Law incentives for juvenile recruiting by drug trafficking gangs: empirical evidence from Rio de Janeiro"
2020/11, Garcia-López, M.A.; Pasidis, I.; Viladecans-Marsal, E.: "Congestion in highways when tolls and railroads matter: evidence from European cities"
2020/12, Ferraresi, M.; Mazzanti, M.; Mazzarano, M.; Rizzo, L.; Secomandi, R.: "Political cycles and yardstick competition in the recycling of waste. evidence from Italian provinces"
2020/13, Beigelman, M.; Vall Castelló, J.: "COVID-19 and help-seeking behavior for intimate partner violence victims"
2020/14, Martínez-Mazza, R.: "Mom, Dad: I'm staying" initial labor market conditions, housing markets, and welfare"
2020/15, Agrawal, D.; Foremny, D.; Martínez-Toledano, C.: "*Paraísos fiscales*, wealth taxation, and mobility"
2020/16, Garcia-Pérez, J.I.; Serrano-Alarcón, M.; Vall Castelló, J.: "Long-term unemployment subsidies and middle-age disadvantaged workers' health"

2021

2021/01, Rusteholz, G.; Mediavilla, M.; Pires, L.: "Impact of bullying on academic performance. A case study for the community of Madrid"

2021/02, Amuedo-Dorantes, C.; Rivera-Garrido, N.; Vall Castelló, J.: "Reforming the provision of cross-border medical care evidence from Spain"
2021/03, Domínguez, M.: "Sweeping up gangs: The effects of tough-on-crime policies from a network approach"
2021/04, Arenas, A.; Calsamiglia, C.; Loviglio, A.: "What is at stake without high-stakes exams? Students' evaluation and admission to college at the time of COVID-19"
2021/05, Armijos Bravo, G.; Vall Castelló, J.: "Terrorist attacks, Islamophobia and newborns'health"
2021/06, Asensio, J.; Matas, A.: "The impact of 'competition for the market'regulatory designs on intercity bus prices"
2021/07, Boffa, F.; Cavalcanti, F.; Piolatto, A.: "Ignorance is bliss: voter education and alignment in distributive politics"

2022

2022/01, Montolio, D.; Piolatto, A.; Salvadori, L.: "Financing public education when altruistic agents have retirement concerns"
2022/02, Jofre-Monseny, J.; Martínez-Mazza, R.; Segú, M.: "Effectiveness and supply effects of high-coverage rent control policies"
2022/03, Arenas, A.; Gortazar, L.: "Learning loss one year after school closures: evidence from the Basque Country"
2022/04, Tassinari, F.: "Low emission zones and traffic congestion: evidence from Madrid Central"
2022/05, Cervini-Plá, M.; Tomàs, M.; Vázquez-Grenno, J.: "Public transportation, fare policies and tax salience"
2022/06, Fernández-Baldor Laporta, P.: "The short-term impact of the minimum wage on employment: Evidence from Spain"
2022/07, Foremny, D.; Sorribas-Navarro, P.; Vall Castelló, J.: "Income insecurity and mental health in pandemic times"
2022/08, Garcia-López, M.A.; Viladecans-Marsal, E.: "The role of historic amenities in shaping cities"
2022/09, Cheshire, P. C., Hilber, C. A. L., Montebruno, P., Sanchis-Guarner, R.: "(IN)convenient stores? What do policies pushing stores to town centres actually do?"
2022/10, Sanchis-Guarner, R.: "Decomposing the impact of immigration on house prices"

2023

2023/01, Garrouste, M., Lafourcade, M.: "Place-based policies: Opportunity for deprived schools or zone-and-shame effect?"
2023/02, Durán-Cabré, J.M., Esteller-Moré A., Rizzo L., Secomandi, R.: "Fiscal Knowledge and its Impact on Revealed MWTP in COVID times: Evidence from Survey Data"
2023/03,Esteller-Moré A., Galmarini U.: "Optimal tax administration responses to fake mobility and underreporting"
2023/04, Armijos Bravo, G., Vall Castelló, J.: "Job competition in civil servant public examinations and sick leave behavior"
2023/05, Buitrago-Mora, D., Garcia-López, M.A.: "Real estate prices and land use regulations: Evidence from the law of heights in Bogotá"
2023/06, Rodriguez-Planas, N., Secor, A.: "College Students' Social Capital and their Perceptions of Local and National Cohesion"
2023/07, Obaco, M., Davi-Arderius D., Pontarollo, N.: "Spillover Effects and Regional Determinants in the Ecuadorian Clean-Cooking Program: A Spatiotemporal Econometric Analysis"
2023/08, Durán-Cabré, J.M., Esteller-Moré, A., Rizzo, L., Secomandi, R.: "Has Covid Vaccination Success Increased our Marginal Willingness to Pay Taxes?"
2023/09, Borrella-Mas, M.A., Millán-Quijano, J., Terskaya, A.: "How do Labels and Vouchers Shape Unconditional Cash Transfers? Experimental Evidence from Georgia"
2023/10, Messina, J., Sanz-de-Galdeano, A., Terskaya, A.: "Birds of a Feather Earn Together. Gender and Peer Effects at the Workplace"
2023/11, Pelegrín, A., Vidal, Ll., González, I.: "Diversifying Economic Risks: Japan's Economic Hedging Towards China"
2023/12, Rodríguez-Planas, N., Secor, A., De Balanzó Joue, R.: "Resilience-thinking Training for College Students: Evidence from a Randomized Trial"
2023/13, Arenas, A., Calsamiglia, C.: "Gender differences in high-stakes performance and college admission policies"

2024

2024/01, Wald, G., Cohen, F., Kahn, V.: "Making Jobs out of the Energy Transition: Evidence from the French Energy Efficiency Obligations Scheme"
2024/02, Durán-Cabré, J. M., Esteller-Moré, A., Montolio, D., Vázquez-Grenno, J.: "Can Teachers Influence Student Perceptions and Preferences? Experimental Evidence from a Taxation Course"

ieb@ub.edu
www.ieb.edu

Public Policies

IEB Working Paper