EDUCATION ECONOMICS
https://doi.org/10.1080/09645292.2022.2113860

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

The novel liquid learning system and the online gap in academic performance

Rodrigo Alegría, Pablo Cárabe, Alejandro Chahoud and Ainara González de San Román

IE University, Madrid, Spain

ABSTRACT
This research considers the innovative educational strategy known as the *liquid learning system*, which allows students attending classes either online or face-to-face. This system was implemented for the first time at a private European university in 2020 as a reaction to the Covid-19 pandemic. Emphasis is placed on the effect of the online choice on student academic performance. Using Instrumental Variables to control for self-selection bias, our findings show a significant gap in the form of lower grades for online students. Quantile regressions reveal that those in the lower tail of the grade distribution are the most adversely affected.

ARTICLE HISTORY
Received 6 October 2021
Accepted 8 August 2022

KEYWORDS
Liquid learning; student performance; COVID-19; online gap; instrumental variables

JEL
I20; I21; I23; I24

1. Introduction

Since the spring of 2020, the COVID-19 pandemic has been disrupting life in almost every country in the world, causing an unprecedented crisis in all sectors of our society. In the field of higher education, the current health crisis has accelerated the rise of many disruptive forces that are reshaping the way higher education institutions carry out their missions, producing dramatic transformations in many different aspects. One of the most striking changes has been an increasing move towards a variety of diverse remote learning modalities, changing the way higher education is provided. Even if these distance-learning tools (*Edutech* industry) were already being used in higher education, social distancing restrictions have fostered and magnified this change. In addition to other effects of the COVID-19 pandemic on university students, the sudden switch to non-traditional academic settings and classrooms may have significantly affected students' learning processes and, consequently, their academic performance.

This paper provides novel evidence on the effects of online classes as compared to face-to-face attendance on university students' academic achievement under the umbrella of a new hybrid educational delivery system implemented in a highly selective private European university. This hybrid system allows students to choose to attend the same synchronous lectures simultaneously, either in person (face-to-face) or online (via zoom) – what we will call the 'liquid classroom' – making it possible to compare the two attendance modes considering the same learning circumstances (e.g. same professor, course material, exams, evaluation criteria, etc.). In other words, direct comparison is possible in this instructional setting since it fosters a liquid teaching and learning context for both face-to-face and online students participating in identical class activities in real time. Thus, after controlling for observable student, course, and classroom characteristics and solving the issue of the

CONTACT Ainara González de San Román agod@faculty.ie.edu
Supplemental data for this article can be accessed online at https://doi.org/10.1080/09645292.2022.2113860

attendance mode selected by students, the only fundamental difference is the physical distance of each student within the same liquid classroom.

Even if this hybrid approach implemented in the fall term of 2020 in response to the prolonged restrictions resulting from the COVID-19 pandemic has been successful in continuing to provide higher education services regardless of students' physical location, we hypothesize that there is a gap in academic performance between online students and their counterparts attending class in person. Moreover, our study provides an important contribution to understanding how distance-learning technologies can potentially exacerbate inequalities. As evidence of this, we explore possible heterogeneous effects of the impact of online attendance on academic performance. We believe that our findings also have important long-term implications for the effective design of education policies.

Theoretically, several underlying mechanisms may be at work that could explain this potential online penalty. The importance of social interactions (Taylor and Simard 1975) and the level of students' commitment and engagement (Pascarella and Terenzini 2005), which influence academic achievement, may be negatively affected by the lack of physical interaction for online students that are not on campus or in the classroom. Another inconvenience for online students is the difficulty of managing interactions in a liquid classroom in order to reduce the perceived pedagogical distance (Moore 1993) and increase the degree of presence (Anderson et al. 2000). Additional obstacles for online students include technical problems, missing out on the richness of nonverbal communication and effective monitoring, instructors failing to properly approach online students, potential multitasking, zoom fatigue (Bailenson 2021), among other factors.[1]

There is already a great deal of literature regarding the possible effects of the COVID-19 pandemic crisis on higher education, most of which focuses on the emergency responses implemented in various countries during the lock-down period in terms of the sudden shift from face-to-face to online distance learning (see Crawford et al. 2020 for a map of the higher education solutions in 20 different countries). Another set of papers, mainly based on surveys, analyze the immediate impacts of the COVID-19 health crisis on various aspects of higher education students: mobility, stress, experiences, and perceptions (Aucejo et al. 2020), socio-economic disparities in higher education (Kuhfeld et al. 2020) and inequalities and student well-being (Rodríguez-Planas 2021). A few other papers are limited to analyzing the effects of the COVID-19 crisis on academic outcomes by comparing academic performance in the spring of 2020 and the pre-pandemic semesters (Orlov et al. 2021; Said and Refaat 2021; Gonzalez et al. 2020). Our paper stands out from the previous research by attempting to shed light on the long-term consequences of new educational strategies arising from the pandemic (in particular, the hybrid learning system), which will surely remain in higher education institutions in the long-term rather than merely being implemented as a short-term emergency response.

Regarding comparisons of delivery modes in higher education, numerous studies attempt to determine whether distance learning delivery modes are better than their traditional counterparts in terms of educational outcomes. It can be argued that this pre-pandemic literature is circumstantial and highly context-dependent, providing mixed results (Escueta et al. 2017; Means et al. 2010). Furthermore, most of these studies compare delivery systems that are either fully online or hybrid versus fully face-to-face. It is important to note that the courses analyzed for this research are already part of a hybrid system, where the appropriate comparison is between students' attendance modes rather than delivery modes. This helps us reduce many other sources of variation that could affect students' academic performance and is therefore more effective in isolating the impact of physical distance.

The main empirical problem with the above literature when attempting to provide plausible causal estimates lies in the issue of self-selection or endogeneity in the choice of delivery system. Although a few studies use traditional regression models (Sussman and Dutter 2010; Brown and Liedholm 2002), they are unable to bypass this issue since the choice is clearly not random, resulting in biased and inconsistent coefficient estimates from education production functions, which consequently leads to incorrect inferences. Two different empirical strategies have been implemented in the existing literature to control for this endogeneity problem in order to provide plausible causal estimates. First, many studies employ experimental settings in which students are randomly assigned to different delivery sections of the same course or courses (Figlio, Rush, and Yin 2013; Joyce et al. 2015; Alpert, Couch, and Harmon 2016; Arias, Swinton, and Anderson 2018 and Cacault et al. 2021). A few others use instrumental variable techniques to mitigate this issue of endogeneity, where the most frequently used instrument to control for this potential bias is the commuting time from a student's place of residence to the university (Coates et al. 2004; Xu and Jaggars 2013 and Bettinger et al. 2017). The overall findings of these studies conclude that taking an online course significantly diminishes student performance.

Our empirical strategy to identify the causal online penalty also suffers from a clear problem of endogeneity. In this study, the issue arises from the design of the system itself: students decide whether to attend classes online or in person, which means that the choice is not random, but rather quite the opposite. As rational individuals, we would expect students to evaluate the costs and benefits of the available options and make their decision based on their specific characteristics and personal circumstances. In this analysis, we use the ***distance*** from each student's permanent address to campus as an instrument to predict the student's likelihood of attending online rather than in person. We argue that distance could play a relevant role in the choice to attend class online, especially considering the context of the pandemic (for example, travel and social distance restrictions imposed in different countries) and the international profile of the students at the university in question. Larger values for this instrument are considered to imply greater associated difficulty and cost of face-to-face attendance.

To summarize our findings, based on a sample of 403 students in two different courses distributed in 16 liquid classrooms, we found that online students suffered a significant penalty in their final grade using this new hybrid delivery system, even when controlling for variables such as gender, cohort, nationality, ability, and attendance. On average, online students perform around 0.8 standard deviations (SD) below face-to-face students in our final specification using the IV empirical strategy. The significance of this result is consistent across several robustness checks that we performed. We believe that this study has important policy implications. For instance, universities should pay attention to the academic performance penalties for the least-skilled students in order to redesign and possibly test these new learning delivery systems. Professors should be trained to ensure the same treatment for all students, regardless of the chosen mode of attendance.

The rest of this paper is organized as follows: Section 2 describes the institutional framework of the university in which this quasi-experimental research was performed; Section 3 presents the data used in the empirical analysis and a descriptive analysis; Section 4 includes empirical findings from the estimation of the different regression models with particular emphasis on the robustness of the results; and, finally, the last section presents the conclusions and a summary of our findings.

2. Institutional and quasi-experimental setup

IE University was founded in 2009 as a Spanish private university, owned by Instituto de Empresa S.L. The university offers programs to undergraduate students (approximately 7,000 students are currently enrolled), which are taught in English with a clear, founded international and entrepreneurial approach. For the purposes of this paper, it is important to address the many unique features of IE University in order to contextualize the environment in which this research took place.

First, it is a highly selective university with high tuition costs, which consequently filters the pool of applicants to include students whose families are mostly on the upper end of the income spectrum. Although the university is in Spain, less than 25% of the students are Spanish. Most are international students that come from all over the world, representing over 130 different nationalities. Students are quasi-randomly assigned to non-overlapping groups in each class, making an effort to ensure a relatively balanced distribution of gender/nationalities in each group. Additionally, the attendance policy is quite restrictive, only allowing students to miss 30% of the sessions for any given course. Professors are also assigned to the different groups at random by the course coordinator. Despite having different professors for each group, course coordinators ensure that they all have a similar syllabus, which implies that all students enrolled in the same course should be learning the same content, regardless of their group. Furthermore, the same grading criteria apply to all courses, irrespective of the groups, and all groups have the same final exam for each course. The idea is to ensure homogeneity and fairness in the learning content and the overall study experience.

Finally, IE University has unique approach to dealing with the COVID-19 pandemic. Even before lockdowns began in Spain, IE University had a robust IT infrastructure system in place, which gave the school a comparative advantage when it came time to transition to fully online classes (when lockdown began in March of 2020). From March 2020 until the end of the academic year in June 2020, university courses were held completely online to comply with lockdown procedures. For the 2020–2021 academic year, IE University implemented a new *liquid learning* educational system, flowing freely between the physical and digital worlds, which is intended to be maintained in the long term and not only as an emergency response to the pandemic crisis.

This educational delivery approach mainly consists of a hybrid system. The objective of this dual format is twofold: first, to provide the greatest degree of flexibility to students facing an uncertain environment in their native countries due to the COVID-19 pandemic; and secondly, to create a new delivery format that could be maintained in the long run, independently of the 2020 health crisis. The idea was to enrich the student learning experience by combining live and online learning features and enhancing the acquisition of new highly demanded skills in this changing world. We now turn to explain the hybrid system in more detail.

First, given the above new educational system, IE University trained and supported the faculty in designing and teaching high-quality liquid courses to increase the effectiveness of this new communication in class and fight against potential obstacles. Specifically, during summer 2020, after the COVID-19 lock-down, instructors were trained in both: (1) to correctly use the new IT infrastructure put into place in the classroom (e.g. camaras, screens and zoom) and (2) to effectively use new online teaching methodologies (such as Feedback Fruits, Trello, or Miro) to adapt lectures, course material and assessment methods (Mentimeter or Socrative) to the new system.

Regarding course structure and class organization, the system design requires 70% of the classes to be synchronous (live, either online or in person) and the other 30% to be asynchronous (recorded/not live, available online). For the synchronous classes, professors were physically present in the classrooms, which had been adapted with touch screens, cameras, and microphones. Students had to choose[2] either to attend these synchronous lectures in person or online (via zoom). The synchronous lectures were taught simultaneously to both groups, which is where the term *liquid* comes from; that is, all students, regardless of their choice of attendance mode, have access to the same course material, perform the same asynchronous activities, take the same live lectures in real time with the same professor, have the same classroom peers throughout the entire course and take the same exams with the same evaluation system. Moreover, both tutoring (academic office hours) and student-support services through the program office, were offered to all students in an online format. Exams were given online (via videoconference) to ensure fairness and students were required to keep their cameras on while taking the exams for their professors to monitor possible unethical behavior and minimize cheating. Summarizing, the access to all university resources was sufficiently similar regardless of the attendance mode.

Finally, to avoid potential negative effects on students' performance based on time zone differences (e.g. having classes in the middle of the night), 'morning', 'midday' and 'afternoon' groups were created to avoid the inconveniences of the different time zones where the online students were located.[3] Thus, the new quasi-random assignation of students to non-overlapping groups in each class attempted to ensure a balance distribution of gender, nationalities and in person/online students within an appropriate schedule independently of the location of the students.

Therefore, once we accounted for the obvious problem of attendance mode selection, all of the aforementioned factors allowed us to conduct a unique study to determine the impact of online vs. face-to-face attendance on student performance. To our knowledge, this is the first paper that has succeeded in completely isolating the impact of attendance modes on academic performance (e.g. physical distance) if compared to prior studies that analyze the differences in delivery modes (e.g. online versus traditional courses).

3. Data and descriptive statistics

3.1. Sample description

We will focus on the courses held in the fall term of the 2020–21 academic year (from September to December 2020), which is when the *hybrid learning* strategy was first implemented. We selected two different cohorts of students in the Bachelor in Business Administration (BBA) program[4], which is important to note in order to effectively test whether the effects vary at different stages of higher education.[5] For the sake of simplicity, we will refer to the first cohort as 'first-years' and the second cohort as 'third-years.' Given the impossibility of comparing the same course for these two different cohorts, we have selected the courses that are the most similar in terms of the number of credits, level of difficulty, and evaluation criteria, in addition to being quantitative[6] courses (see complete list of core courses in Table 1A in the Appendix). The two selected courses are: (i) **Applied Business Math, or ABM** (fall term-first year) and (ii) **Data Analysis for Economics, or DAE** (fall term-third year).

We used cluster sampling to select approximately half of the population in each course, where the clusters are the different class groups. Eight groups were randomly selected within each course and then the information was collected from all the students in the selected clusters. We assume that all of the clusters/groups are equally representative of the population of interest (quasi-random assignment). Groups are similar in size at approximately 25 students per group. We excluded the few third-year students that were repeating or retaking a course (approximately 5%) since they were exempt from attending classes and, consequently, most of the relevant information for the analysis would be missing. The final sample includes 403 students distributed in 16 different groups taught by 4 different professors, and the size is balanced between the two cohorts with 192 first-years and 211 third-years. We consider the sample large enough for our purposes when compared to the total size of the BBA program (around 50% of the total number of students).

3.2. Data collection

Our data has been compiled from different sources. First, we collected information from the IE administrative records for all of the students in the sample, specifically gender, cohort, nationality, and permanent residential address. Nationalities were classified into 10 categories by cultural and geographical proximity, ensuring that there was a considerable number of students in each category (see Table 2A in the Appendix for the list of countries in each group). We did not include variables related to family background, such as income or parental education, since the profile of IE students make them rather homogeneous in those characteristics. Next, since students' innate abilities could not be observed, we used a proxy measure of ability considering grades obtained in schooling prior to the university. The majority of research in this field uses high school grade point averages to proxy students' unobserved abilities prior to higher education (Danilowicz-Gösele et al. 2017).[7] In this case, given the major variations in the different education systems of the international students in our sample, we approximate each student's ability with their official certified grade used to access the university[8] also taken from the administrative records. This provides a measure of prior ability, which is comparable across all students, regardless of their country of origin, which is in line with previous studies on academic success (McKenzie and Schweitzer 2001).[9]

Second, we merged individual-level data on student performance by asking all professors to provide detailed information on the grades given throughout the course (class participation, mid-term exam, final exam and final grade), the chosen system of class attendance (online or in person) and the general attendance rate throughout the course, irrespective of whether students were physically or virtually present in the classroom. Finally, more specific information on the two selected courses (ABM, and DAE) was collected for the analysis, including students' respective groups and their corresponding professors.[10] Our intention is to control for the influence of both unobserved (group fixed effects) and observed (professor quality) common factors that simultaneously affect the academic performance of students in the same group.

Regarding the outcome variable to estimate academic performance, we used the final grade for the corresponding course for each student. Since those grades are not comparable between the two cohorts as the courses are different, they have been standardized to zero mean and unit standard deviation[11] and will be called **Standardized Final Grades** for the purposes of this paper. We are aware that grades are an imperfect measure of what students may actually learn in a course; nevertheless, they provide some indications about learning and, consequently, student success, as pointed out in the relevant literature (e.g. Bettinger et al. (2017).

Regarding the main explanatory variable, which is central to our analysis, we created the **online** dummy variable as a binary indicator with a value of 1 for students that chose online attendance (as reported by their professors) and 0 for students that chose to attend class in person. There are two noteworthy aspects inherent to this variable. First, we assume that the choice of online or face-to-face attendance is constant throughout the semester; that is, each student's selected attendance mode is considered to be invariant from the beginning of the course straight through to the end. Considering the COVID-19 pandemic situation and the major differences between students' nationalities and cultures in our sample, we believe that this is a reasonable assumption (which is also supported by the professors' experiences), especially for students that opted for online classes. It is indeed very likely that the majority did not return from their places of residence after the summer of 2020 due to the health crisis and the restrictions imposed in various countries around the world. On the other hand, we are aware that there may be more variability for students whose attendance status was reported as face-to-face, since they could also connect online if classes were very early in the morning or due to poor weather conditions, for example.[12] However, if we consider the central hypothesis of this research – the existence of an online attendance penalty – then relaxing this assumption would lead to underestimating the online penalty, thereby obtaining a lower bound of the impact, which we consider sufficiently credible.

Second, the attendance choice is not exogenous in this context (the allocation is not random, but rather depends on the individual). Students were likely to choose one of the two options

Table 1. Variable Information: Type, Nomenclature and General Description.

Type of Variable (**Name**)	Variable Description
Student Performance	
Std. Final Grade	Standardized Final Grade (mean 0 and unit standard deviation)
Student Attendance	
Online	=1 if student attended classes online, 0 for face-to-face
Attendance	Attendance rate for the course (percentage from 0 to 100)
Student Prior Ability	
Ability	University entrance grade (homogenized on a scale of 0-10)
Student Demographics	
Female	=1 if student is female, 0 if student is male
First-year	=1 if student is in the first cohort, 0 for the second cohort
Nationalities	Dummies for the 10 different groups of nationalities
Groups and Instructors	
GrABM1-GrABM8	Dummies for the 8 different groups first-years may belong to
GrDAE1-GrDAE8	Dummies for the 8 different groups third-years may belong to
InstABM1-InstABM2	Dummies for the 2 different professors teaching ABM
InstDAE1-InstDAE2	Dummies for the 2 different professors teaching DAE

given their specific characteristics or circumstances during the pandemic, meaning that the differences between the two groups could be a result of the systematic sorting of students in each of the attendance modes. The next section discusses our empirical strategy to overcome this problem of endogeneity. Table 1 provides a description of all the variables discussed in this section.

3.3. Descriptive analysis

Table 2 presents a summary of the students' characteristics for the entire sample, classified by attendance mode. The last column shows the difference between the two and its significance according to the t-test of two independent samples.

As shown in Table 2, about 40% of the students in our sample are females. Attendance levels are usually very high[13], with an average of approximately 95% of the total number of sessions. The mean level of *ability* in the pooled sample is around 8.114, with a small deviation of 0.818. This shows how demanding the university is in terms of entrance grades and confirms the homogeneity of students in terms of prior ability.

Interestingly, for the purposes of this research, we have extracted valuable information regarding the distribution of students' characteristics when comparing the two attendance systems. First, online students attend classes relatively and significantly less than face-to-face students (nearly a 4% difference), which may indicate that the latter are slightly more motivated.[14] Second, there is a significantly higher proportion of first-years among the group of face-to-face students (55.6%). We argue that the uncertainty of newcomers may increase their need to attend classes in person as compared to more experienced students. Third, it seems that the choice of how to attend classes is not conditioned by gender, as the percentage of females is significantly similar in both groups. However, face-to-face students are on average more skilled.

Figure 1A in the Appendix presents the frequencies of the different nationalities in each system, revealing that there is a higher proportion of Spanish, French and Italian students and a lower proportion of Latin American and Asian students attending class in person. It seems evident that returning to Spain in the middle of a pandemic could be much more expensive and riskier for students whose place of residence is further away, which consequently increases the incentives to choose the online system.

The evidence presented in Table 2 and Figure 1A (the attendance choice clearly depends on certain traits of each student, such as nationality, ability and motivation) confirms the endogeneity problem that arises when trying to identify the causal online penalty and justifies the use of an instrumental variable approach to solve it.

Next, we focus on the average effect of online attendance (versus face-to-face) on academic achievement by cohort. A simple comparison of mean grades between the two groups of students

Table 2. Average Student Characteristics: Online versus Face-to-face.

	All	Online	Face-to-face	Difference*
Female	0.406	0.402	0.414	0.012
	(0.492)	(0.491)	(0.494)	(0.049)
First-year	0.476	0.418	0.556	0.137***
	(0.500)	(0.494)	(0.498)	(0.050)
Attendance	95.175	93.503	97.491	3.988***
	(7.366)	(8.633)	(4.144)	(0.717)
Ability	8.114	7.978	8.302	0.324***
	(0.818)	(0.781)	(0.833)	(0.081)
Sample size	403	234	169	

Notes: Standard deviations in parentheses. *Statistical significance tested using two-sample *t*-tests assuming independent samples and equal variances. Significance levels are indicated in the last column by *** $p < 0.01$. Groups and professors are not included in the table since all groups are similar in size and each professor taught half of the groups in each course.

Table 3. Raw Gap of Online Attendance on Academic Performance by Cohort.

Standardized Final Grade	Online	Face-to-face	Raw Gap*
All	−0.248	0.344	−0.592***
	(1.055)	(0.798)	(0.096)
Sample Size	234	169	
First-years	−0.261	0.272	−0.533***
	(1.105)	(0.796)	(0.139)
Sample Size	98	94	
Third-years	−0.239	0.434	−0.673***
	(1.022)	(0.797)	(0.136)
Sample Size	136	75	

Note: Standard Deviations in parentheses; *Comparison of means t-test with independent samples and equal variances. Significance levels are indicated in the last column by *** $p < 0.01$.

serves to quantify the raw gap. Table 3 shows the means and standard deviations for the Standardized Final Grade for students attending face-to-face and online by cohort. Column 4 of this table quantifies the difference in means of the two systems of class attendance. The descriptive information in Table 3 reveals that raw gaps are negative in all cases and of considerable size. Specifically, online students (58.06% of the sample students), on average, perform worse than their peers who are physically present in the classroom. In particular, online students obtain an average Std. Final Grade that is 0.592 standard deviations (SDs) lower. If we look at the differences by cohort, we can see that online first-years score an average of 0.533 SDs less than their peers who attended class in person. Furthermore, the penalty is greater (in absolute value) for third-years for which the online penalty is 0.673. In both cases, the statistic for the means comparison confirms that the difference is statistically significant.

In addition to computing the raw differences of the mean, it is also very important to examine the entire grade distribution in order to determine to what extent they vary between online and face-to-face students. Figure 1 depicts kernel density estimates of Std. Final Grades by selected attendance system and cohort. The dashed and solid lines correspond to online and face-to-face students, respectively. Panel (a) shows the comparison for first-years and panel (b) shows the corresponding comparison for third-years.

Observing Figure 1, we can clearly appreciate that the distribution of grades for online students is located further to the left in both cases although the difference is more remarkable for third-years than for first-years, confirming the empirical evidence found in Table 3. We conducted the Kolmogorov–Smirnov test for equality of distributions[15] and reject the null for both cohorts, statistically

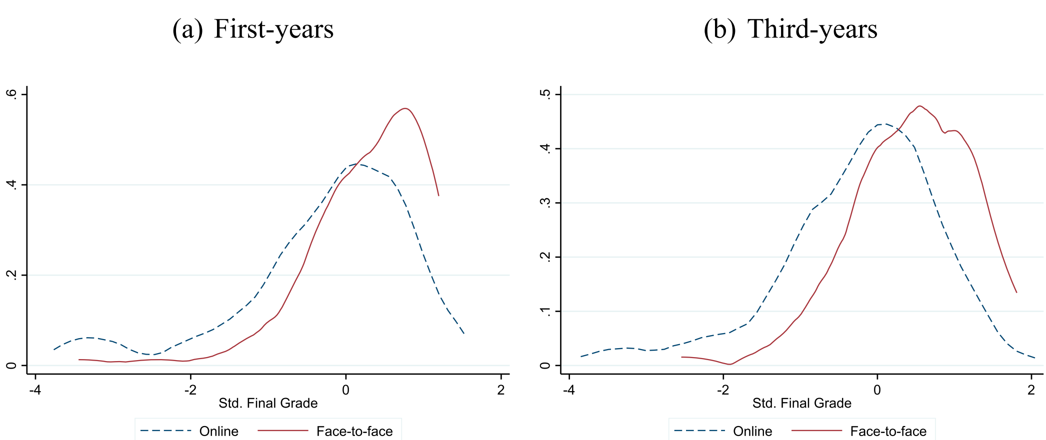

Figure 1. Density Std. Final Grade by selected system (Online/Face-to-face) and cohort.

confirming that the distributions cannot be considered equal in either of them. The figure also suggests that the effect may be different depending on the area of the distribution where the student is located, showing a greater impact for students with lower grades. We analyze those distributional aspects in more detail as a robustness check in Section 4.

In summary, the descriptive exercise in this section allows us to identify remarkable differences in academic achievement between online and face-to-face students, both at the mean and throughout the grade distribution. These raw differences are more pronounced for third-years. However, we are cautious with this initial evaluation since these gaps could be partially attributed to the different compositions of students in the two different attendance modes, therefore requiring a more formal analysis. The following section presents our empirical strategies and the results obtained.

4. Identifying the impact of online attendance on academic performance

This section presents the main results of the analysis. First, we describe the empirical approach that supports the results. Second, we validate the empirical evidence from the descriptive analysis through the estimation of different econometric models using Ordinary Least Squares (OLS) and Instrumental Variables (IV) estimation strategies. Finally, we perform several robustness checks to validate the results.

4.1. The empirical approach

To quantify the impact of online attendance on students' academic performance, we consider an equation for student *i* in group *k* using the theory of education production functions[16] as follows:

$$y_{ik} = \gamma Online_i + \beta X_i + \mu_i + \delta_k + \varepsilon_{ik} \tag{1}$$

where y_{ik} is the academic performance of student *i* (measured with the Standardized Final Grade); *Online* is a dummy variable for the student attending classes online instead of in person; *X* is a set of observed characteristics of the student (such as gender, nationality, cohort, prior ability,[17] and attendance) associated with a vector of coefficients β; μ_i represents the unobserved characteristics of student *i*; δ_k represents the student's academic return due belonging to group *k* (such as peer effects, class composition and professor quality effects). Finally, ε_{ik}, is the error term, which is assumed to follow a normal distribution.

The main focus of this study is to estimate the impact of online class attendance on academic performance, γ. The key question is to find the most suitable estimation strategy for this context. If μ_i and δ_k were uncorrelated with $Online_i$, which is conditional on the observed student and group characteristics, γ can be consistently estimated by OLS applied to Equation (1). However, such estimates may suffer selection biases resulting from the omission of unobserved variables correlated with the chosen system of class attendance. If this is the case, the estimation of γ by OLS will include the true effect as well as the bias factors arising from student and group heterogeneity.

Accordingly, the relevant question when estimating the online effect is how to eliminate these biases. In order to control for unobserved heterogeneity across groups we can introduce Group and Instructor Fixed Effects into the model. However, controlling for student unobserved heterogeneity is a more challenging task. To deal with this issue, we have employed an IV approach, using a variable related to the treatment (online), but theoretically unrelated to the outcome (academic performance). In this analysis, we use the ***distance*** from each student's permanent address to campus as an instrument to determine the student's likelihood of attending online rather than in person. Specifically, we first identified the permanent address associated with each student in the dataset (City, Zip Code, Country) as registered in the university's internal records and then used Google Maps to calculate the distance as a straight line (in *km*) between each student's place of residence and the IE University campus (Madrid).[18]

There is one major concern about using distance as an instrument in this context. Some researchers have indicated that families who value education may choose to live closer to campus (e.g. Card 1995; Long and Kurlaender 2009). The specificities of our sample of international students from high-income families may limit the impact of this issue, as they are very unlikely to change their permanent address (assumed to be their family residence) for this reason. Logically, students whose places of origin are very far from the university are likely to live in an apartment near campus; however, given the COVID-19 pandemic circumstances and the timeframe of our research, even if this were the case, we believe that most of them would have returned to their family residence during the summer of 2020.

Returning to the technical details, we estimate Equation (1) together with the following first-stage equation, in a two-stage least squares (2SLS) framework:

$$Online_i = \alpha_0 \, log(distance_i) + \alpha_1 X_i + \delta_k + v_{ik} \tag{2}$$

where *Online* is replaced in Equation (1) by the predicted values resulting from the probit model that estimates Equation (2). Given that $distance_i$ is heavily skewed to the right, ranging from less than 1 km to 12,190 km in our data (see Figure 2A in the Appendix), we take the natural logarithm in the first-stage to improve the model's fit. This estimation strategy would imply that the estimated coefficient γ from Equation (1) implemented using *distance* represents an unbiased estimate of the online penalty, but only if distance is indeed an appropriate instrument, which we will check below.

4.2. Empirical estimation of the online penalty: OLS and IV strategy

Table 4 presents a series of regression models for the entire sample of students where the dependent variable is the **Standardized Final Grade**. Two different estimations are shown: **(a)** OLS estimates and **(b)** IV estimates. The first model includes only the dummy variable *Online*, which takes on the value of one if the student attended classes online. Model (2) incorporates observed student characteristics such as gender, nationality, attendance, the proxy for ability, the corresponding cohort, and

Table 4. Estimated Impact of Online Attendance on Student's Standardized Final Grades.

	(a) OLS Estimates			(b) IV Estimates – log(distance)		
	Model (1)	Model (2)	Model (3)	Model (1)	Model (2)	Model (3)
Online	−0.592***	−0.421***	−0.418***	−0.588***	−0.803***	−0.746***
	(0.105)	(0.121)	(0.117)	(0.236)	(0.255)	(0.287)
Female		0.212***	0.169**		0.209***	0.168**
		(0.078)	(0.087)		(0.078)	(0.087)
Attendance		0.045***	0.045***		0.040***	0.045***
		(0.008)	(0.008)		(0.010)	(0.008)
Ability		0.407***	0.401***		0.379***	0.387***
		(0.068)	(0.066)		(0.064)	(0.065)
First-year		−0.357***	−0.838***		−0.529***	−0.492***
		(0.113)	(0.141)		(0.195)	(0.192)
First-year × Online		0.153	0.258		0.391	0.537
		(0.172)	(0.183)		(0.367)	(0.351)
const	0.344***	−7.331***	−6.959***	0.343***	−6.371***	−6.646***
	(0.072)	(0.731)	(0.797)	(0.140)	(1.205)	(1.283)
Nationalities	×	Yes	Yes	×	Yes	Yes
Group FE	×	×	Yes	×	×	Yes
Instructor FE	×	×	Yes	×	×	Yes
Sample Size	403	403	403	403	403	403
Adj.R^2	0.085	0.321	0.333	0.013	0.313	0.326

Notes: Standard errors clustered by group in parentheses. Nationalities incorporate to the model dummies for the different grouped nationalities – 9 in total (+ reference). Group FE incorporate dummies for the different groups in both courses – 15 in total (+ reference). Instructor FE incorporate dummies for the different teachers (one per course + reference). The estimates are reported as being significant at * $p < 0.1$, ** $p < 0.05$ and *** $p < 0.01$.

the interaction between the cohort and the *Online* dummy. The last model also includes Fixed Effects (FE) for the different groups and professors in the sample.

Considering the results from panel (a), Model (1) results in a high penalty – online attendance reduces students' final grades by 0.6 SDs, and the difference is statistically significant. Clearly, this simple regression does not allow for comparing students with similar characteristics and, consequently, the penalty is equivalent to the raw gaps analyzed in Section 3. Adding student observed characteristics, Model (2), increases the explanatory power of the model and results in a reduction of the online penalty, suggesting that part of the gap is due to the systematic sorting of students between the two different systems of class attendance, as we saw in the descriptive analysis, but it still represents a major, significant impact. Adding group and instructor fixed effects in Model (3) barely changes the results of the previous model, suggesting that the online gap is robust across groups and professors. The online penalty from the IV estimates in Panel (b) is still negative and statistically significant across different model specifications and higher in absolute value as compared to the traditional estimation strategy. In particular, the IV estimate of Model (3), our preferred specification, suggests that the online penalty is as big as 0.746 SDs. The above comparison suggests that OLS significantly underestimates the online penalty in our study, which is in line with the findings of Coates et al. (2004) or Xu and Jaggars (2013).

Note that the credibility of these findings hinges on the validity of the instrument used in the estimation. We have therefore performed two tests to assess the validity of our IV. The first is to test whether the IV is a weak instrument, as the IV's lack of a strong predictive power for the endogenous variable may bias the IV estimate toward the OLS estimate (Bound, Jaeger, and Baker 1995; Greene 2018). As reported in Table 5, the first-stage regression results indicate that permanent address distance indeed has a highly significant predictive power for students' choice to attend online, even after netting out the impacts of other included explanatory variables such as gender, ability, cohort, nationalities, and group and professor fixed effects in Model (3). Consistently, formal IV tests yield F-statistics that are larger than 50, which are substantially greater than the widely used threshold[19], providing little evidence of a potentially weak IV.[20]

Another potential concern is that the instrument may have a direct effect on the outcome variable. Although we only used one source of exogenous variation in students' attendance status to achieve identification, there is a natural set of other IVs to facilitate an over-identification test. As suggested by Nevo and Rosen (2012), under the assumption that the IV used is uncorrelated with the error term ε in Equation (1), any functions of this IV are also uncorrelated with ε and can therefore serve as additional IVs. Following this suggestion, we used both *distance* and *distance squared* as IVs for *Online* and re-estimated our preferred model (Model (3) in Table 4). The new IV estimate of the online penalty ($\hat{\gamma} = -0.673$, se = 0.268) is very similar to its counterpart in column (3) of Table 4. The results of the over-identification test (Sargan Chi2 = 0.629, p-value = 0.427) provide a robustness check of our results to alternative functional forms of the IV. Other functions of the *distance* variable (e.g. *distance* alone, and dummies for 0-50 km, 50-1000 km, 1000–5000 km, and >5000 km) were also tried; the IV estimates remained quite similar.[21]

Table 5. Results of First-Stage IV – Probability of Online Attendance.

	Online		
Dep.Variable	Model (1)	Model (2)	Model (3)
log(distance)	0.201***	0.317***	0.343***
	(0.026)	(0.048)	(0.054)
F-statistic	69.81	62.93	55.34
Pseudo R^2	0.114	0.249	0.314

Notes: Standard errors clustered by group in parentheses. Each column shows the first-stage coefficient from a probit estimation where the dependent variable is the Online dummy. The different model specifications correspond to those in Table 4. The second row shows the F-statistic for the excluded instrument from the first-stage. The estimates are reported as being significant at *** $p < 0.01$.

It is still possible that the entire variation extracted by these IVs is correlated with the error term ε in Equation (1), and that standard over-identification tests failed to detect this problem. We therefore performed a falsification test to assess the validity of our IV. Following Xu and Jaggars (2013), we tested whether the IV for the treatment (e.g. Online in this case) has a significant predictive power for the outcome variable for the non-treated (e.g. face-to-face students), who should presumably not be affected. The estimated coefficient for the IV is very small and insignificant ($\hat{\gamma} = -0.114$, se = 0.346), suggesting that even if distance does have some direct effect on the outcome variable, this effect would not overturn the finding of a statistically significant online penalty.

Additionally, one could argue that the main source of variation for the instrument is the difference between international and domestic students given that the median distance in our data is 3000 km. Although the admission criteria to the university is the same no matter the origin of the students, we also checked using our data whether the domestic students resemble international students in terms of observed ability. For this purpose, we first defined a binary indicator for domestic students which equals one for those students whose place of residence is less than 800 km from campus.[22] The proportion of domestic students in the sample is 31.02%. Next, we computed the average observed ability for the two groups and found a value of 8.08 for international students while a value of 7.98 for domestic ones, being the difference statistically insignificant at 1% level. Nonetheless, if differences in unobserved ability were to exist, the nature of our cross-sectional data does not allow to untangle it.

We also checked whether differences in the observed characteristics are correlated with distance. Table 3A in the Appendix presents the Pearson correlation matrix of the instrument with the observed characteristics used in this paper, except for the grouped nationalities (which will obviously correlate with distance). None are significantly correlated with distance (correlations no larger than 0.06 in absolute value) except for the endogenous covariate, which reinforces the previous conclusion of distance being a good predictor of the online choice (the correlation is 0.388). Additionally, we conducted an alternative test by creating a *predicted outcome* variable using the OLS Model (3) in Table 4 and computing its correlation with the instrument (distance). The correlation is −0.07 which suggests that there is no relationship between course outcomes and distance. This evidence of independence strengthens the validity of our IV. In any case, we controlled for all observed characteristics in the preferred specifications to minimize confounding effects.

The last concern we addressed in relation to the validity of the instrument, given the broad range of distance among students in our setting, is whether results from the IV strategy (more importantly, in the first-stage estimation) remain unchanged if we restrict the sample to students living within commuting distance. We first defined *Commuting Distance* as a binary indicator being equal one for students who lived less than 30 km from campus.[23] And secondly, we replicated the IV estimation of Model (3) in Table 4 and get consistent results in both stages. The coefficient of log(distance) barely changes with respect to the baseline first-stage with all the observations but is statistically significant only at 10% level (the standard error goes up to 0.201 due to the reduced sample size). This evidence suggests that the IV results were not driven only by students living very far away.

The estimated coefficients of other explanatory variables also provide pertinent information. Female students tend to perform better on average than their male counterparts, while first-years tend to perform worse on average than third-years. Attendance is a crucial determinant of success in quantitative courses – students attending more classes (regardless of the selected system) do get higher scores on average. Ability turns out to be individually significant, which is in line with the literature citing similar proxies for ability (McKenzie and Schweitzer 2001). Furthermore, the online penalty is lower for first-years, given that the interaction term is positive and, therefore, compatible with the evidence provided in the descriptive analysis. However, that interaction appears to be insignificant even in the most sophisticated specification, suggesting that there is no significant difference between the online penalty for the two cohorts once all the observed characteristics have been controlled.

4.3. Robustness checks

This section presents three robustness checks to address relevant questions raised in the previous sections and test the sensitivity of our results. Specifically, we re-estimated our preferred specification (Model (3) in Table 4) with the two estimation strategies (OLS and IV), changing various characteristics of the original model or sample. First, as the final grade has been identified to be a less-controlled measure of performance than the mid-term and final exams (Joyce et al. 2015), we replaced the dependent variable with other measures of the course grade (panel A of Table 6), such as participation grade, mid-term, final exam grades, and the probability of failing the course. Second, we tried to determine whether the penalty associated with online attendance is different by skill groups. To address this question, we divided the final grade distribution into percentiles (10th, 25th, 75th and 90th) and then estimated separated models for each of them. Panel B summarizes the estimated online penalties for the different quantiles. (See also Figure 3A in the Appendix for a graphical introspection). Finally, Panel C shows the test of whether our proxy for ability can be considered good enough for third-year students. We consequently only reestimated the model for the sample of third-years, where the proxy for ability was replaced by (i) the student's final grade in the Statistics course (taken during the fall of 2019), which may be a better proxy for the contemporaneous skills in quantitative courses, and (ii) the student's accumulated GPA until the fall term of 2020, which could be considered a more stable measure of ability.z

Several interesting features emerge from the robustness checks presented in Table 6. The penalty is robust for the use of different dependent variables as shown in Panel A. The negative effect is found to be greater for the participation grade than for the exam items of the evaluation system. Although these results may confirm the difficulties faced by online students participating in the liquid classroom, they could also indicate differences in the professors' approaches to the new learning system and their rather weak ability to effectively encourage online student involvement in this setting.[24]

Table 6. Robustness Checks.

	Participation	Mid-term Exam	Final Exam	Probability of Failure
Panel A. Dependent Variable				
OLS	−0.601***	−0.390***	−0.371***	0.515**
	(0.089)	(0.102)	(0.106)	(0.238)
IV	−1.112***	−0.786***	−0.465**	0.425
	(0.262)	(0.222)	(0.256)	(0.651)
Panel B. Heterogeneity by Skill Group – Quantile Regression				
	10th	**25th**	**75th**	**90th**
OLS	−0.760**	−0.523***	−0.410***	−0.311***
	(0.324)	(0.192)	(0.149)	(0.113)
IV	−1.029***	−0.854***	−0.562**	−0.441
	(0.389)	(0.274)	(0.247)	(0.309)
Panel C. The Ability Measure – Third-years				
	Baseline Ability	**Statistics Grade**	**Accumulated GPA**	
OLS	−0.356***	−0.362***	−0.182*	
	(0.115)	(0.088)	(0.088)	
IV	−0.903**	−0.727**	−0.604**	
	(0.368)	(0.345)	(0.246)	

Notes: The columns in each panel contain the coefficient estimate of the Online dummy from a separate regression using both OLS and IV strategies. The specification is from the most sophisticated model in Table 4, however, changes were made to the estimation between each panel. In Panel A (N = 403), the dependent variable was replaced with different course grades; in column 4 the dependent variable is an indicator = 1 if the student failed the course (final grade < 5); columns 1–3 use the continuous standardized grades described in the panel labels. Panel B (N = 403) uses quantile regression for the different percentiles of the final grade distribution. Panel C (N = 211) uses different ability measures for the restricted sample of third-years with column 1 as the baseline. Standard errors in parentheses are clustered in groups. The estimates are reported as significant at * $p < 0.1$, ** $p < 0.05$ and *** $p < 0.01$

The choice to attend classes online also affects the probability of failing the course. Moreover, the coefficients of the probit model presented in the last column of Panel A cannot be interpreted as probability changes (they do not represent the strength of the effect), but still indicate its direction and significance. Nevertheless, we can easily compute the marginal effect of online attendance on the probability of failure by using the probit model equations where the values for the rest of the explanatory variables are set at their mean levels. In particular, the OLS estimation yields a significant difference of approximately 6%, which disappears when the IV strategy is used, probably due to inefficient and inflated standard errors in the 2SLS estimation, given the small proportion of failures in our sample (only 10.92% of students failed their class).

Second, if we observe the estimates reported in Panel B, we see that the online penalty decreases (in absolute value) as we move towards the right-hand side of the grade distribution. In the most sophisticated specification, the gap goes from [−1.029, −0.760] in the bottom quantile (10th) to [−0.441, −0.311] in the top quantile (90^{th}). This evidence suggests that the online penalty is greater for students in the lower tail of the grade distribution, producing a potential increase in inequality between low-skilled and high-skilled students, as has been found in previous research (e.g. Bettinger et al. 2017 and Cacault et al. 2021). It could be argued that students at the upper tail of the grade distribution might find it easier to understand the course material on their own, meaning that they benefit from self-efficiency in study habits. Moreover, highly skilled students may find it easier to participate in the liquid classroom even if they are attending online since their participation is likely to provide a quality contribution to the discussion. On the other hand, unprepared students may not feel capable of making such contributions, which is aggravated by their lack of interaction with the professor when attending online. In short, these results suggest that the new hybrid system could be increasing inequalities among students, although it also has a positive counterpart – that good students are less affected therefore, having a lesser effect on university excellence. Seen from another perspective, this also raises the question of whether potential cheating among low-skill students (which is easier a priori given the online nature of the exams) leads to any benefit in the form of higher grades.

Finally, the robustness check in Panel C validates the proxy for ability used throughout this paper. In particular, the online penalty is similar and significant across the different measures of ability, suggesting that we can consider our proxy strong enough to control for innate ability. Another remarkable finding that arose from this test is the huge similarity between the baseline results and the results obtained using the Statistics grade. This may indicate that, even if our original ability measure could be considered a weak proxy for current third-years' ability, it is encapsulating quantitative skills better than the accumulated GPA.

5. Conclusion

This paper presents evidence of a new hybrid learning system, which allows university students to freely decide whether to attend classes online or in person within the post lock-down context of the COVID-19 pandemic. To our knowledge, this is the first study of its kind that provides evidence of the online penalty in a hybrid educational system. Our quasi-experimental design allows for the direct comparison between online and face-to-face students, as the two groups both participate in identical class activities in real time. The physical distance within the same liquid classroom is considered the only fundamental difference between the two groups.

Based on a sample of 403 students, with the corresponding data regarding their specific characteristics and academic outcomes, we have examined the impact of the choice to attend online classes on academic performance in a highly selective private university.

Our results show that students choosing to attend lectures online experienced a penalty in the form of lower grades, which ranges from 0.4–0.8 SDs across various model specifications. This relationship is significant even when using an IV approach to address the issue of self-selection for the online option. Furthermore, the online penalty documented in this paper presents heterogeneity by skill group: students at the lower tail of the grade distribution are the most adversely affected, to a significant degree.

Our findings indicate that online attendance cannot entirely replace or substitute attending class in person. Disadvantages such as lower levels of participation, engagement and student interaction, in addition to a lack of the professors' experience dealing with this new hybrid system are a few factors that explain the poorer academic outcomes for online students as compared to face-to-face students.

Nevertheless, several factors may influence the generalizability of these results to other educational settings. First, the gaps we observed at IE University may be even wider in universities with less academically prepared and socially disadvantaged students. Second, some universities may be more effective than others in terms of how they design and support liquid lectures. IE University has already expended its considerable resources to support the faculty in designing and teaching high-quality liquid courses. Specifically, the university trained instructors after the COVID-19 lock-down to adapt to the IT infrastructure put into place in the classroom, as well as providing ongoing support throughout the semester. Finally, the sample design does not allow for the extrapolation of results to all courses taught in the university. It may be possible that the effects are diminished in less quantitative courses as students are more capable of working on their own and are not so professor-dependent.

Although we cannot test for the possible mechanisms driving the effects, we believe this study has important policy implications. Given the heterogeneity of the penalties documented in this paper, higher education institutions should be especially careful with less-advantaged students to redesign and test these new learning delivery systems.

Further research is clearly needed to fully understand this new liquid learning system and its implications on students' academic outcomes. We would have liked to collect more information on other courses and not only quantitative courses to test potential differences arising due to the nature of the course. Although we are confident that our findings are relevant to improving the implementation strategies for these new learning methodologies in higher education, we also believe that more research is necessary to empirically identify the mechanisms through which differences in academic achievement may arise based on the attendance mode.

Notes

1. Di Pietro et al. (2020) provides a detailed literature review of the main channels through which the adoption of distance education may negatively influence students' learning across all educational levels.
2. There could be several reasons behind a student's choice to attend online or in person based on their personal circumstances such as income (saving transportation and accommodation costs), geographical distance, aversion to risk due to COVID-19, ability, motivation, etc.
3. In August 2020 the university asked all the students what their place of residence would be during the academic year precisely to guarantee that the time zone would not hinder their class attendance. More specifically, students within Madrid time and +/- 1 h were allocated in the 'morning', 'midday' and 'afternoon' groups. In the 'midday' groups, students located in +/-4 h respect to Madrid time were also added. Students whose residence during the course was -5 h or more were assigned to the 'afternoon' groups. And finally, students living +5 or more were added into the 'morning' groups.
4. The BBA program students represent approximately 70% of the total number of the undergraduates.
5. Given the main objectives of this paper, we decided to use first and third-year students as they represent the two extremes of age and experience. We did not include fourth-year students since the fall term of the third year is the last term of compulsory courses; from that point on, students choose elective courses and the assigned groups per se cease to exist.
6. We chose quantitative courses as we expect the effect of the selected class attendance system (online or face-to-face) to be much more evident than in less demanding courses.
7. Using a unique administrative data set from a German university, the authors show that high school grades are strongly associated with graduation probabilities and final grades, whereas variables measuring social origin or income have a smaller impact.
8. This is the university entrance score taken from various official certificates, depending on the student's nationality, which converts grades from different educational systems to the Spanish system. These grades are taken from the credentials issued by UNED (*Universidad Nacional de Educación a Distancia*) in order for foreigners to

access the Spanish university. In the case of Spanish students, the grade is taken from the EVAU (*Evaluación para el Acceso a la Universidad*) results.

9. The authors examined a group of approximately 200 first-year students and found significant coefficients for the university entrance score, which account for 39% of the variance in GPA.
10. There are two professors for both courses, each teaching half of the groups in the sample.
11. This standardization ensures that grades follow the same 'scale' for both courses/cohorts in order to make them comparable.
12. See Cacault et al. (2021) for a detailed analysis of the increased use of live streaming classes associated with shocks to health or commuting conditions.
13. Attendance policy in the university is rather restrictive, requiring students to attend a minimum 70% of the total number of sessions in the course in order to pass.
14. We argue that attendance may be considered a good proxy of student motivation.
15. The two-sample Kolmogorov-Smirnov test compares the cumulative distributions of two data sets. The null hypothesis is that both groups were sampled from populations with identical distributions.
16. The relationship between empirical specifications of education production functions and the underlying theory is examined in Todd and Wolpin (2003) and Hanushek (1979), among others.
17. Prior ability is approximated with each student's Official Certified Grade to access the university, as explained in Section 3.
18. Distance varies substantially among the various students. The median student lives 3000 km from campus, and the interquartile range is 433–5686 km. Approximately 25% of the students in the sample reported their permanent residence at more than 5000 km away from campus, making face-to-face attendance quite unlikely.
19. Stock, Wright, and Yogo (2002) describe a rule of thumb for estimating the strength of the instrument in models using one instrumental variable for an endogenous covariate, as in the case of this study. The variable is considered to be a weak instrument if the F-statistic against the null hypothesis is less than 10.
20. We tried with another IV that is closely related to the pandemic: the COVID-19 Stringency Index, which is a composite measure at the country level based on nine response indicators including school closures, workplace closures and travel bans, rescaled to a value from 0 to 100 (100 = strictest) from *Our World in Data Org* (Oxford); however, the IV turned out to be weak.
21. The estimate obtained using *distance* as the IV for online, without taking log, is -0.681 SDs (of the distribution of final grades) and the estimate obtained using a set of distance dummies as described in the text is -0.662 SDs, both of which are significant at the 1% level. Detailed results are available upon request.
22. We define Domestic using distance to campus rather than nationality since at IE University there are Spanish students living abroad and foreign students living in Spain, so that we believe is a more accurate indicator for the domestic students. We take a threshold value of 800 km in the definition since it is the maximum distance between any geographical location in Spain with its capital city (Madrid).
23. We assume 30 km as a reasonable maximum commuting distance for college students living in Madrid taking into account the pandemic situation in which mobility was very limited. In our sample, the proportion of students living within commuting distance is 18.61% (75 observations), of which 30% attended classes online.
24. Introducing interactions of professor dummies with the Online indicator in the specification from Panel A reveals that this could be one of the mechanisms behind the participation penalty since two of them appear to be statistically significant. Results are available upon request.

Acknowledgements

This paper has benefited from the help and support of many individuals from the IE University administration and faculty. Our gratitude also goes out to seminar participants at IE University Brownbag (May 2021) for their helpful comments and suggestions. We would like to thank Francisco Machin Aragones, Isabel de Sivatte, Giovanna Lamastra, Samuel Mancebo, Maud Pindard-Lejarraga, Daniel Fernandez-Kranz, Patricia Gabaldon, and Stephanie Lackner. The authors also appreciate the anonymous referees' valuable comments and suggestions. Any errors that remain are our sole responsibility. E-mail: ainara.gonzalez@ie.edu; ralegria@faculty.ie.edu

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Alpert, William T., Kenneth A. Couch, and Oskar R. Harmon . 2016. "A Randomized Assessment of Online Learning." *American Economic Review* 106 (5): 378–382.

Anderson, Terry, D. Randy Garrison, Walter Archer, and Liam Rourke. 2000. “Methodological Issues in the Content Analysis of Computer Conference Transcripts.” *International Journal of Artificial Intelligence in Education* 12: 8–22.

Arias, J. J., John Swinton, and Kay Anderson. 2018. “Online vs. Face-to-Face: A Comparison of Student Outcomes with *Random* Assignment.” *E-Journal of Business Education and Scholarship of Teaching* 12 (2): 1–23.

Aucejo, Esteban M., Jacob French, Maria Paola Ugalde Araya, and Basit Zafar. 2020. “The Impact of COVID-19 on Student Experiences and Expectations: Evidence from a Survey.” *Journal of Public Economics* 191: 104271.

Bailenson, Jeremy N. 2021. “Nonverbal Overload: A Theoretical Argument for the Causes of Zoom Fatigue.” *Technology, Mind, and Behavior* 2 (1): 1–6.

Bettinger, Eric P., Lindsay Fox, Susanna Loeb, and Eric S. Taylor. 2017. “Virtual Classrooms: How Online College Courses Affect Student Success.” *American Economic Review* 107 (9): 2855–2875.

Bound, John, David A. Jaeger, and Regina M. Baker. 1995. “Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak.” *Journal of the American Statistical Association* 90 (430): 443–450.

Brown, Byron W., and Carl E. Liedholm. 2002. “Can web Courses Replace the Classroom in Principles of *Microeconomics*?” *American Economic Review* 92 (2): 444–448.

Cacault, M. Paula, Christian Hildebrand, Jérémy Laurent-Lucchetti, and Michele Pellizzari. 2021. “Distance Learning in Higher Education: Evidence from a Randomized Experiment.” *Journal of the European Economic Association* 19 (4): 2322–2372.

Card, David. 1995. “Using Geographic Variation in College Proximity to Estimate the Return to Schooling.” In *Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp*, edited by L. N. Christofides, E. K. Grant, and R. Swidinsky, 201–222. Toronto: University of Toronto Press.

Coates, D., B. R. Humphreys, J. Kane, and M. A. Vachris. 2004. ““No Significant Distance” Between Face-to-Face and Online Instruction: Evidence from Principles of Economics.” *Economics of Education Review* 23 (5): 533–546.

Crawford, Joseph, Kerryn Butler-Henderson, Jürgen Rudolph, Bashar Malkawi, Matt Glowatz, Rob Burton, Paulo Magni, and Sophia Lam. 2020. “COVID-19: 20 Countries’ Higher Education Intra-Period Digital Pedagogy Responses.” *Journal of Applied Learning & Teaching* 3 (1): 1–20.

Danilowicz-Gösele, Kamila, Katharina Lerche, Johannes Meya, and Robert Schwager. 2017. “Determinants of Students’ Success at University.” *Education Economics* 25 (5): 513–532.

Di Pietro, Giorgio, Federico Biagi, Patricia Costa, Zbigniew Karpiński, and Jacopo Mazza. 2020. *The Likely Impact of COVID-19 on Education: Reflections Based on the Existing Literature and Recent International Datasets*. Luxembourg: Publications Office of the European Union.

Escueta, M., V. Quan, A. J. Nickow, and P. Oreopoulos. 2017. “Education Technology: An Evidence-Based Review.” *NBER Working Paper No. 23744*.

Figlio, David, Mark Rush, and Lu Yin. 2013. “Is it Live or is it Internet? Experimental Estimates of the Effects of Online Instruction on Student Learning.” *Journal of Labor Economics* 31 (4): 763–784.

Gonzalez, Teresa, M. A. De La Rubia, Kajetan Piotr Hincz, M. Comas-Lopez, Laia Subirats, Santi Fort, and G. M. Sacha. 2020. “Influence of COVID-19 Confinement on Students’ Performance in *Higher* Education.” *PloS one* 15 (10): e0239490.

Greene, William H. 2018. “Econometric analysis/Limdep users manual.”.

Hanushek, Eric A. 1979. “Conceptual and Empirical Issues in the Estimation of Educational Production Functions.” *Journal of Human Resources* 14 (3): 351–388.

Joyce, Theodore J., Sean Crockett, David A. Jaeger, Onur Altindag, and Stephen D. O’Connell. 2015. “Does Classroom Time Matter?.” *Economics of Education Review* 46: 64–77.

Kuhfeld, Megan, James Soland, Beth Tarasawa, Angela Johnson, Erik Ruzek, and Jing Liu. 2020. “Projecting the Potential Impact of COVID-19 School Closures on Academic Achievement.” *Educational Researcher* 49 (8): 549–565.

Long, Bridget Terry, and Michal Kurlaender. 2009. “ Do Community Colleges Provide a Viable Pathway to a Baccalaureate Degree?” *Educational Evaluation and Policy Analysis* 31 (1): 30–53.

McKenzie, Kirsten, and Robert Schweitzer. 2001. “Who Succeeds at University? Factors Predicting Academic Performance in First Year Australian University Students.” *Higher Education Research & Development* 20 (1): 21–33.

Means, Barbara, Yuki Toyama, Robert Murphy, Marianne Bakia, and Karla Jones. 2010. *Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation, and Policy Development.

Moore, Michael G. 1993. “Three Types of Interaction.” In *Distance Education: New Perspectives*, edited by K. Harry, M. John, and D. Keegan, 19–24. London: Routledge.

Nevo, Aviv, and Adam M. Rosen. 2012. “Identification with Imperfect Instruments.” *Review of Economics and Statistics* 94 (3): 659–671.

Orlov, George, Douglas McKee, James Berry, Austin Boyle, Thomas DiCiccio, Tyler Ransom, Alex Rees-Jones, and Jörg Stoye. 2021. “Learning During the Covid-19 Pandemic: It is not who you Teach, but how you Teach.” *Economics Letters* 202: 109812. doi:10.1016/j.econlet.2021.109812.

Pascarella, Ernest T., and Patrick T. Terenzini. 2005. “How College Affects Students: A Third Decade of Research.” *Volume 2*. Jossey-Bass, An Imprint of Wiley. 10475 Crosspoint Blvd, Indianapolis, IN 46256.

Rodríguez-Planas, Núria. 2021. “COVID-19 and College Academic Performance: A Longitudinal *Analysis*.” *IZA Discussion Paper No. 14113*.
Said, El, and Ghada Refaat. 2021. “How Did the COVID-19 Pandemic Affect Higher Education Learning Experience? An Empirical Investigation of Learners’ Academic Performance at a University in a Developing Country.” *Advances in Human-Computer Interaction* 2021: 1–10.
Stock, James H., Jonathan H. Wright, and Motohiro Yogo. 2002. “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments.” *Journal of Business & Economic Statistics* 20 (4): 518–529.
Sussman, Stephen, and Lee Dutter. 2010. “Comparing Student Learning Outcomes in Face-to-Face and Online *Course* Delivery.” *Online Journal of Distance Learning Administration* 13 (4): 6–11.
Taylor, Donald M., and Lise M. Simard. 1975. “Social Interaction in a Bilingual Setting.” *Canadian Psychological Review/ Psychologie Canadienne* 16 (4): 240–254.
Todd, Petra E., and Kenneth I. Wolpin. 2003. “On the Specification and Estimation of the Production Function for Cognitive Achievement.” *The Economic Journal* 113 (485): F3–F33.
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.” *Economics of Education Review* 37: 46–57.