Taking advantage of COVID-19? Online learning, descentralization and tertiary education

Elisa Failache

INSTITUTO DE ECONOMÍA
Serie Documentos de Trabajo

Agosto, 2023
DT 09/2023

ISSN: 1510-9305 (en papel)
ISSN: 1688-5090 (en línea)

Este trabajo se basa en un capítulo de mi tesis doctoral defendida en noviembre 2022 en la Universidad Autónoma de Barcelona. Quiero agradecer especialmente a Xavier Ramos, Almudena Sevilla, Gabriel Facchini y Natalia Nollenberger. Además quiero agradecer por los comentarios recibidos en la conferencia ECINEQ. Por ultimo, quiero agradecer al Ministerio de Ciencia e Innovación de España por el apoyo para la realización de mi doctorado mediante la beca de Formación de Personal Investigador (BES-2017-080211).

Forma de citación sugerida para este documento: Failache, E. (2023) "Taking advantage of COVID-19? Online learning, descentralization and tertiary education". Serie Documentos de Trabajo, DT 09/2023. Instituto de Economía, Facultad de Ciencias Económicas y Administración, Universidad de la República, Uruguay.

Taking advantage of COVID-19? Online learning, descentralization and tertiary education

Elisa Failache*

Abstract

Distance to university has been shown as an important factor for students' decisions to continue studying after secondary school and for their academic outcomes. Therefore, the generalized shift to online learning could have opened a window of opportunity for those students living in places lacking a university campus. Following a difference-in-differences strategy, I compare students that already lived in an area with or close to the university (control group) with students that live far away (treated). I take advantage of the institutional setting in Uruguay, where the public university is free and without entrance exam but with campuses only in half of the territory. The data come from administrative records for the period 2017-2021. Results show less adverse effects in terms of dropouts for the treated freshmen students of 2020 but no effects on other academic outcomes conditional on dropout. In addition, I find an enrollment increase in localities without university campuses for 2021.

Keywords: online learning, educational outcomes, internal migration, difference-in-differences

JEL Classification: I22, I23, I24

(*) Elisa Failache, IECON, Universidad de la República, Uruguay, correo electrónico: elisa.failache@fcea.edu.uy

Resumen

La distancia a la universidad es un factor importante para las decisiones de los estudiantes de continuar estudiando después de la secundaria y para sus resultados académicos. Por lo tanto, el cambio generalizado al aprendizaje en línea derivado de la pandemia podría haber abierto una ventana de oportunidad para aquellos estudiantes que viven en lugares lejanos a la universidad. Para analizar esto aprovecho el marco institucional de Uruguay, donde la universidad pública es gratuita y sin examen de ingreso pero con campus solo en la mitad del territorio y sigo una estrategia de diferencias en diferencias, comparando estudiantes que ya vivían en un área cerca de la universidad (grupo de control) con estudiantes que viven lejos (tratados). Los resultados muestran menos efectos adversos en términos de deserción para los estudiantes de primer año tratados de 2020, pero ningún efecto en otros resultados académicos condicionados a la deserción. Además, encuentro un aumento de la matrícula en localidades sin campus universitarios para el 2021.

Palabras claves: aprendizaje en línea, resultados educativos, migración interna, diferencia en diferencias.

Código JEL: I22, I23, I24

2 Introduction

There is a common consensus that higher education is crucial to promote growth and development, not only benefiting the individual but society as a whole.[1]. While for developed economies the tertiary enrollment rate is more than 75%, for developing countries the figure is less favorable, being 38% for middle-income countries.[2] Therefore, analyzing policies to promote tertiary education enrollment, particularly for developing countries, is imperative. In addition, the COVID-19 pandemic and the suspension of face-to-face lessons posed several challenges to the educational system. In most countries, online learning was the response to the impossibility of teaching courses in situ. Tertiary education was not an exception. As many papers have shown, this solution may have widened educational gaps as there could be uneven access to online resources, physical space, or an adequate environment for learning (Rodríguez-Planas 2022*b*, Bacher-Hicks et al. 2021). On the other hand, access to virtual learning could reduce the costs associated with studying for students living far away from the university. The distance to the place of residence has been shown as an important factor in students' decisions on attending an educational center and in their academic outcomes (Alm & Winters 2009, Frenette 2009, Lapid 2016). Therefore, the generalized shift to online learning could have opened a window of opportunity for those students living in places where supply of tertiary education is nonexistent. In this chapter, I analyze this hypothesis.

As mentioned, COVID-19 triggered the shift to online learning, implying a reduction of the distance to university virtually to zero. I take advantage of this shift in a particular institutional setting to analyze if the pandemic, and the subsequent shift, affected individuals living far away from a university campus differently in terms of (i) the academic outcomes, for already enrolled students and (ii) the enrollment decision, for potential students. I exploit the particular institutional setting of Uruguay, which is advantageous for four reasons. First, the main public university (Universidad de la República - UDELAR), which covers 85% of tertiary students in the country, has campuses only in half of the territory. Campuses are located in 8 out of 19 departments, and the largest educational offer is concentrated in the capital city, Montevideo. Therefore, after finishing high school, many students who aim to continue studying have no choice but to move to another city.[3] Second, this university is free of tuition and without entrance exams, ruling out other possible causes as discouraging elements of attending university. Third, the pandemic broke out after the 2020 enrollment, implying that enrollment decisions and course registration were already decided for 2020 cohort of students. By comparing students' academic outcomes in 2020 to previous years, I can measure the effect of the pandemic and the online learning shift (i) on academic outcomes. Fourth, in contrast to 2020, in 2021, new students could enroll and register for courses knowing that classes were going to be online. By comparing

[1]See for example United Nations Educational, Scientific and Cultural Organization (https://www.unesco.org/en/education/higher-education/need-know) or World Bank Education Overview Caño-Guiral (2018)

[2]The gross enrollment ratio is defined as total enrollment over the population of the age group that officially corresponds to the level of education shown. For Latin American countries, the figure is 54%, while for OECD countries, it is 77%, and 87% for the US. Figures obtained from the World Bank Dataset (https://data.worldbank.org/indicator/SE.TER.ENRR) for the year 2019.

[3]In 2019, more than 25 % of new students moved to another city

enrollment rates by geographical localities in 2021 to previous years, I can measure the effect (ii) on enrollment decisions.[4]

I use a rich dataset obtained from different administrative record sources from UDELAR, which contains information on first-year enrolled undergraduate students from 2017 to 2021. In particular, these administrative records have information on students' performance at university and their sociodemographic and socioeconomic characteristics. The empirical strategy follows a difference-in-differences strategy. First, I define the treated group as those students from localities far away from the university campus. For the treated group, COVID-19 and the subsequent switch to online learning implied the possibility of return to (or avoiding leaving) their hometowns and/or reducing commuting long distances. The fact of having this new possibility is what I call treatment. I compare treated and control freshman students' academic outcomes in 2020 versus their peers enrolled in previous years in which face-to-face classes prevailed. Second, I aggregate the number of freshmen students at the locality level and compute the enrollment rate by localities. I compare treated localities (those without a university campus) with control localities (those with a university campus) in 2021 and before.

Results show that due to the pandemic, there was a general increase in university dropout rates.[5] However, dropout was less pronounced for treated students, that is, for students who could move back to their towns or avoid commuting long distances due to the shift to online learning. I do not find differential effects of treatment in other academic outcomes. In addition, the effect was slightly more pronounced for girls. In terms of enrollment rates, I find an increase in the enrollment rate by locality in those localities without a university campus. The size of the effect implied an increase of 13% compared to the levels before 2021 for treated localities, suggesting that online learning could be a strategy to increase tertiary education enrollment. All results are robust to different specifications.

This paper relates to two strands of the literature. On the one hand, I contribute to the literature that analyses the role of distance in access to tertiary education. Several papers show the importance of distance in the decision to continue studying, career choices, and academic outcomes (Alm & Winters 2009, Spiess & Wrohlich 2010, among others). However, a challenge in this literature is associated with the cofounders of the student's decision. Frenette (2009) and Lapid (2016) contribute to this problem using the expansion of universities to get the causal effect of distance. In addition, this literature is more scarce in developing countries. An exception is a paper by Katzkowicz et al. (2021) that also exploits the expansion of the university campuses in Uruguay to measure the effects on enrollment. Overall, studies suggest that distance is a relevant factor in understanding students' academic decisions and outcomes. However, the role of online learning and its effect as a way of reducing distance is still an open question. With this study, I contribute to improving this gap.

On the other hand, several papers analyzed the effect of COVID-19 on different

[4]According to the National Institute of Statistics, geographical localities (or census localities) are defined in terms of clearly and precisely delimited territories made up of clusters of buildings, and therefore reflect the representation of landscape changes. A Census Locality corresponds to a set of census tracts characterized by a concentration of population and dwellings. I present a description of these localities in the Appendix A

[5]I define dropout by the fact of having enrolled in the university but not doing any academic activity after

outcomes for developed economies finding: positive effects on dropout rates (Aucejo et al. (2020), Rodríguez-Planas (2022*b*)); delay in graduation ((Aucejo et al. 2020), Rodríguez-Planas (2022*b*)); improvements in GPA (Rodríguez-Planas (2022*a*), Bulman & Fairlie (2022)); or no effect on academic outcomes (Bonaccolto-Topfer & Castagnetti (2021)). In addition, some papers focused on understanding the effects of online learning, triggered by COVID-19, on academic outcomes finding negative effects on grades (Kofoed et al. (2021), Bird et al. (2022), De Paola et al. (2022), Altindag et al. (2021)). For developing economies, the results are more scarce, finding an increase in withdrawal from courses (Jaeger et al. (2021), Failache et al. (2022)), difficulties in access to technology (Hossain (2021), Jaeger et al. (2021)), and positive effects in grade Failache et al. (2022). I contribute to this literature first by analyzing the effects for a Latin American country, therefore, adding to the knowledge for countries outside the US, the main country analyzed. Second, I build on the literature by considering a particular group of students that could have benefitted more from the pandemic: students living far away from a university campus.

The rest of the paper is organized as follows. Section 3 describes the conceptual framework behind the analysis, and section 4 presents the most relevant related literature. The institutional setting is presented in Section 5, followed by the data description in Section 6. In Section 7, I develop the empirical strategy. Finally, Section 8 and Section 9 present the results and final remarks respectively.

3 Conceptual framework

The conceptual framework behind this analysis is based on the idea that distance is relevant in the choices of students. When students finish high school, they face two different decisions. First, they have to define whether to continue studying or not. Second, conditional on continuing studying, they have to choose where to enroll. In both decisions, the distance is likely to matter (Alm & Winters 2009). When there is no tertiary institution near the student's hometown, attending higher education requires migrating or commuting, and this fact could discourage enrollment. In addition, once the decision to attend university is taken, if the student decides to move or commute long distances, this can have consequences on students' outcomes via a more constrained budget and/or a time-consuming activity. In addition, migration could affect students socioemotionally. There could be a positive side regarding more independence or discovering new things, but also a negative side related to a feeling of loneliness or difficulties in adapting to a new place.

Because distance can affect students' academic decisions and outcomes, online learning could affect differentially those students for whom distance is a potential binding restriction. First, online learning could affect the decision to continue studying due to cost reductions. That is, affecting university enrollment. Second, even if that decision had been to attend university, online learning could save time (from commuting and/or adapting to a new place), thus affecting academic outcomes. Therefore, online education could be an opportunity for improving educational outcomes, particularly for students living far away from a university campus.

It is worth mentioning that, besides the channels mentioned before, online learning could have additional effects on students outcomes for several reasons. As De Paola et al. (2022) point out, online learning has benefits and drawbacks compared to face-to-face lessons. On the benefits, when course recordings are available, students can attend classes when they prefer, avoiding too crowded classrooms. In addition, they can review lessons as many times as they want. Regarding the drawbacks, the lack of in-person peer interactions and interactions with professors could negatively affect students. Moreover, technology-related issues such as unreliable internet or difficulties in technological skills may undermine the learning process. Besides, the lack of routines and timetables might induce students to procrastinate, making study more difficult (De Paola et al. 2022). In my setting, both groups (treated and control) are being exposed to online learning. Therefore, if I assume that online learning affects students living closer or far away from the university similarly, the distance is the salient factor explaining the different results in my analysis.

4 Literature review

4.1 Student internal migration

As mentioned before, in many cases, the decision to study at the university goes together with the decision to migrate. The literature on student migration to attend tertiary education mainly focused on the US and involved mostly interstate migration, a small part of total migration. In addition, many of these papers focused on the role of financial aid and tuition in the migration process (Alm & Winters 2009). The papers analyzing the role of distance in enrollment and tertiary educational outcomes are more scarce.

Alm & Winters (2009) study interstate college migration using a gravity model with data from Georgia, finding that the distance from a student's home to the university campus is a relevant variable in the decision. In particular, results show that the probability of attending any tertiary institution decreases with the distance to college elasticity being less pronounced in more prestigious institutions. Spiess & Wrohlich (2010) analyze the role of distance in demand for higher education using data from the German Socio-Economic Panel and university postal codes. They estimate a discrete choice model and find that, after controlling for socio-economic and regional characteristics, the distance to the nearest university affects the enrollment decisions of high-school students. The results suggest that the distance effect is driven mainly by transaction costs rather than by neighborhood effects. However, as Gibbons & Vignoles (2012) point out, one problem of this literature relates to the estimation of causal effects of home-university distances on the decision choice of students due to co-founders driving the results (such as spatial heterogeneity or residential sorting,). The authors try to overcome this issue by using a large administrative dataset for England that can account for many student characteristics and estimate reduced form logit specifications on individual student-level microdata. They find that the geographical distance to the university has little or no impact on the participation decision but is relevant to the institutional choice. Yet, as the authors stated, the distances to the nearest institutions are relatively small in England. In addition, incorporating student fixed effects and a broad set of characteristics could still have endogeneity issues.

To overcome the endogeneity issues, Frenette (2009) exploits the opening of universities in cities in Canada to provide causal evidence of the importance of distance for university and college participation rates. Results show an increase in local youth's university attendance and a reduction in college participation in most cities. Overall, the effect is an increase of 1.3 percentage points in postsecondary participation. These effects are particularly relevant for lower-income family students. Lapid (2016) also exploits the openings of universities to test the importance of distance as a binding constraint for four-year college enrollment. Using data from California, the author uses event study and difference-in-differences models and found a 1.5 percentage points increase in the four-year enrollment rate among recent high school graduates from local high schools. In addition, there is no effect on the share of local graduates who attend farther-away campuses, suggesting minor crowd-out effects compared to impacts on the extensive margin.

This literature is even more scarce for developing countries. Jardim (2020) analyze the impact of university opening on educational outcomes of students using an event study approach with two-way fixed effects. The author finds an average increase of 0.038 SD in test grades in municipalities where the university opened. More related to this work, Katzkowicz et al. (2021) analyze the effect of the expansion of UDELAR outside the county's capital on total enrollment and the share of first-generation university students using a difference-in-differences framework. They find that the decentralization process successfully increased the number of students from localities outside the capital and also increased the share of students with parents that do not hold a university degree.

Overall, the literature that analyses the role of distance in tertiary education suggests that distance is a relevant factor in understanding students' academic decisions and outcomes. However, this literature is still scarce, with most papers providing non-causal evidence. In addition, the role of online learning and its effect as a way of reducing distance is still an open question.

4.2 COVID-19 and tertiary education

The literature about the effects of COVID-19 is broad and addresses multiple dimensions such as labor markets, health, economic growth, inequality and education (Bacher-Hicks et al. 2021, Chetty et al. 2020, among others). Regarding education, many papers analyze the effects of the pandemic on elementary school, high school, and tertiary education. In this section, I focus on the work done on the impacts of COVID-19 on tertiary education outcomes.

Using a survey sample of 1500 students from a university in the US (Arizona State University), Aucejo et al. (2020) analyze the causal impact of the pandemic by using a questionnaire instrument that collects information about what different outcomes/expectations would have been observed in the absence of COVID-19. Results related to academic performance show that COVID-19 affected the delay in graduation by 13%, increased by 11% the students that withdrew from classes and 12% the students intending to change major. In addition, around 50% of the sample reported a decrease in study hours and academic performance. The authors also find a reduction in preferences for online instruction based on the recent experience of students. The effects are heterogeneous according to different characteristics. As an example, the results by socio-economic backgrounds show that low-income students are more likely to postpone the decision to graduate (55%), more affected in their major choice decision (41%), and COVID-19 implied an increase of nearly 100% of the expected Grade Poing Average (GPA) gap increasing inequalities among groups.

Rodríguez-Planas (2022*a*) uses administrative records from a college in New York (Queens College - City University of New York) to identify the effects of the COVID-19 pandemic on academic performance using a difference-in-differences and event study approach with individual fixed effects. The author analyses differences in the impact across lower- and higher-income students, finding that lower-income students outperformed their higher-income students. The result is driven mainly by the lower-income students in the bottom quartile of the Fall 2019 cumulative GPA, that obtain a 9% higher GPA than their higher-income peers. Suggestive evidence supports the idea that this result could be due to challenges with online learning faced by lower-income top-performing students. In addition, the differences in GPA are explained by a flexible grading policy adopted by the university. Besides, Rodríguez-Planas (2022*b*) uses the same dataset and additional information from an online survey collected in 2020 to estimate the causal impact of the pandemic on other academic outcomes. The author finds that the pandemic caused between 14% and 34% of the students to consider dropping a class, a reduction in freshman students' retention rate by 26%, and 30% of students modified their graduation plans, with two-fifths of them postponing graduation. Also using administrative data for the US but for students in the 116-college California Community College system, Bulman & Fairlie (2022) analyze the trend of enrollment, fields of study, and academic outcomes and how these were affected by the pandemic. They found a drop in students enrolled of 11% from 2019 to 2020 and 7% from 2020 to 2021. The reduction was most significant among African-American and Latin students. Regarding academic outcomes, conditional on enrollment, from spring 2019 to spring 2020, course completion fell from 73% to 71%, but course grades of "A" increased from 40% to 50% together with a decrease in grades "B" and "C".

Different results are found by Bonaccolto-Topfer & Castagnetti (2021) using administrative data for an Italian university (University of Pavia). The authors use a difference-in-differences design comparing students' outcomes during the summer term of 2020 to students in the same term but of the previous years and find no substantial effects of COVID-19 on teaching quality and academic performance measured by grades, graduation rates, and exam failure. The results are similar even considering heterogeneous groups according to family wealth, top-performance students or gender.

Because COVID-19 also implied a switch to online learning, some papers focus on understanding the effects of online learning on academic outcomes. Kofoed et al. (2021) analyze the results from a randomized control trial that took place in the fall 2020, where students were assigned either to online or in-person classes for an Introductory Economics course in a US Military Academy. The results show a decrease of 0.215 SD in students' final grades of the ones that took the online course. The authors conducted a survey to disentangle the mechanisms, finding that online students struggled to concentrate in class and felt less connected to their instructors and peers. Bird et al. (2022) use the shift to virtual classes and follow the difference-in-difference framework taking advantage of administrative records of a university in Virginia, US. They estimate a within-instructor-course variation, comparing students that started courses (during Spring) in person or online, and a student fixed effects equation. Both approaches lead to a modest negative effect of online learning, between 3% and 6%, on course completion, driven mainly by an increase in course withdrawals but also by the rise in course failure. Students with lower GPAs suffered more from online teaching. De Paola et al. (2022) also follow the difference-in-differences strategy to investigate the impact produced by the shift on the teaching modality in an Italian university (University of Calabria) using administrative records. The authors compare students' performance in the second semester versus the first semester of 2020 and contrast this with the same difference in previous academic years. Results show adverse effects of online teaching in credits courses per semester (0.11 standard deviations) and for an overall measure of students' performance that considers grades obtained. Results are worst for first-year students. Finally, Altindag et al. (2021) use administrative records from a US public university that already had many online courses before the pandemic and shifted all courses to virtual in the fall of 2020. They estimate a flexible equation that controls for the year and term together with student and instructor fixed effects. Results show that online teaching implied a worse performance in terms of grade, the propensity to withdraw from a course, and approval of the course for students. A relevant finding in their setting is that without the inclusion of instructor-specific factors, the relationship would lead to mistakenly concluding that online classes have better academic outcomes. Once including the fixed effect, face-to-face teaching shows better results for students.

All previous studies focused on developed economies. The literature about the effects of COVID-19 on tertiary education for developing economies is scarce. Hossain (2021) uses survey data from the Young Lives Study, collected in Ethiopia, India, Peru, and Vietnam, to describe differences in the effects of remote schooling according to sociodemographic characteristics. Not surprisingly, using logit regressions, the author finds that students from wealthier households, urban areas, and with internet access are more likely to access remote schooling. In addition, Jaeger et al. (2021) conducted a large worldwide survey to students in many countries, including Mexico, as the only developing economy. Considering respondents from all countries, in terms of educational outcomes, they found that 12% of the students withdrew from at least one course and 41% were not sure about returning to school in the fall of 2020. In addition, 83% of students manifested the lack of contact with faculty or students as a challenge. For Mexico, an additional relevant problem was the lack of a noiseless place to study or lack of access to the internet or computer. Directly related to this study are the results found in Failache et al. (2022) that analyze the effect of COVID-19 on university students in Uruguay using the same administrative records as this paper. The paper estimates the difference in academic outcomes in 2020 compared to previous years. University students in Uruguay dropout more in 2020 and took fewer courses than in previous years. Conversely, the mean grade was higher than in previous years.

My analysis contributes to this literature by understanding the differential effect that the pandemic could have had on a particular group of students: those living far away from the university. For these students, the pandemic and the consequential shift to online learning could be a solution to the distance as a limitation for attending the university.

5 Institutional Setting

5.1 Universidad de la Republica

University educational system in Uruguay is characterized by the concentration of students in Universidad de la Republica, the main public university in the country. UDELAR offers around 100 undergraduate degrees and more than 200 postgraduate degrees and hosts 86% of Uruguayan university students. One distinctive characteristic of UDELAR is that there are no tuition fees nor admission exams, making university education accesible for everyone.[6] [7] However, because the graduation rate in secondary school is low, the gross enrollment tertiary ratio is 65%.[8] In 2019, 140.000 undergraduate students were enrolled at UDELAR, from which close to 20.000 were new students.

The second distinctive characteristic of the Uruguayan tertiary system is its geographical concentration. Uruguay is organized into 19 geographical administrative units, called departments, of which Montevideo is the capital. Located in the south centre of the country, Montevideo is the smallest department in terms of extension but the most populated, with half of the population living there (close to 1.3 million people).[9] Most of UDELAR's supply degrees are offered only in Montevideo, and this is the case also for the majority of courses provided by private universities, vocational or teacher training programs.

Since 2007 a university territorial decentralization process has taken place by progressively expanding the supply of degree programs over the country. By 2017, when the last expansion occurred, seven out of nineteen departments had a university building in their capital city, with 8 degree programs offered on average per department (Figure 1). As an example of the effect of the decentralization policy in terms of distances, the expansion implied that for someone living in Artigas, the department furthest from Montevideo, before the decentralization, the university was 500 km away. After the expansion, Artigas has the closest university campus 130 km away, in Rivera.[10] Despite the decentralization process, the percentage of students enrolled in the campus in Montevideo is still the vast majority, around 85% in 2019, with 56% of students that lived outside the capital the previous year to enter university (Udelar 2020). This implies that for a substantial number of students, commuting for long periods or migrating to the capital is a factor to take into account when deciding to go to the university.

[6] There are a limited number of bachelor degrees for which the access is defined by lottery given the limited number of slots

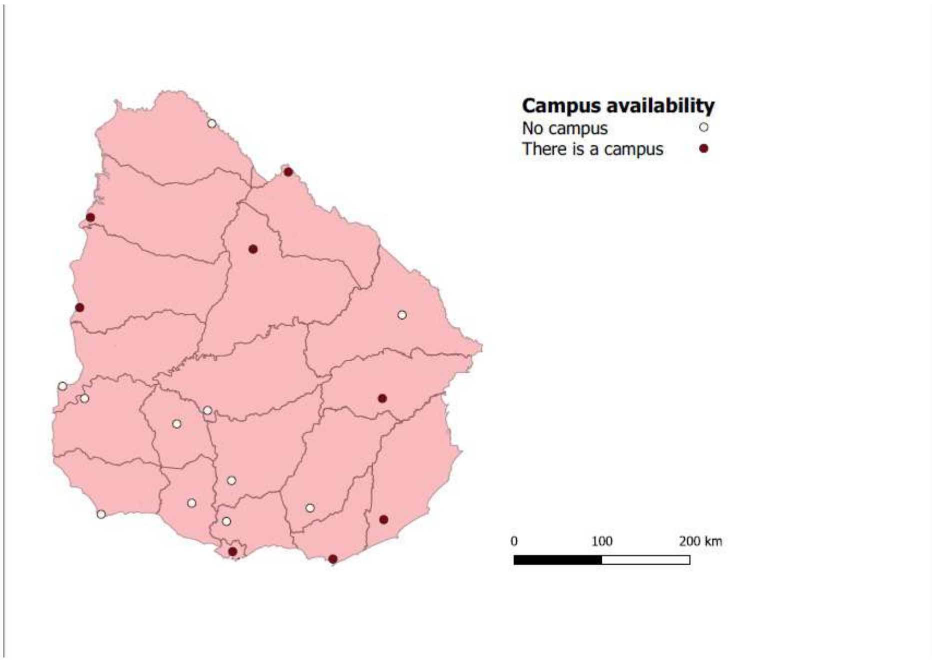
[7] Some postgraduate degrees have tuition fees but the majority are free of charge.

[8] Data obtained from the World Bank dataset (https://data.worldbank.org/indicator/SE.TER.ENRR)

[9] A map with density and total population per department is presented in the Appendix A

[10] The distance is calculated using the minimum distance from one point to the other. Therefore, I could be underestimating the commuting distance and time given that the roads in Uruguay were thought to connect different places with Montevideo, but less to connect other departments between each other

Figure 1: Presence of University by departments in 2017 onwards



Finally, it is worth mentioning that the general enrollment process of first-year students consists of attending the university in person during February, and the only requirement is to have finished high school. After enrollment, students register for those courses they would like to attend within the first year of their degree. Both annual and semestral courses coexist depending on the degree. The academic year starts in March and ends in December.

5.2 COVID-19 and institutional responses

The first COVID-19 patient detected in Uruguay was on the 13th of March of 2020, when courses at the university had just begun. By that date, the pandemic was already causing alarm around the world. Therefore, by mid-March, the university authorities decided to suspend courses for one month at the undergraduate and graduate levels. At the national level, the government did not impose a generalized lockdown in the country at any time, but one day after the university authorities' decisions, they decided to suspend the classes in all the educational system. In addition, teleworking was strongly suggested in private firms and mandatory in public offices.

In UDELAR the suspension of in-person classes continued, and by mid-April, courses started switching to virtual classes. The implementation of online teaching was defined at the course level and was not homogeneous between courses. However, to carry on the virtual learning process, the UDELAR used tools that it had previously developed and incorporated new ones. Specifically, 380 virtual teaching rooms were offered, with a capacity for up to 1000 students to be simultaneously connected and attending lessons. By May 2020, virtual tools were widespread and used in all university degrees. It is worth mentioning that over the last decades, a wide array of policies to foster the ICT sector was implemented in Uruguay to provide high-quality internet connection and guarantee digital inclusion in all the country. As a result, in 2021 86% of the population used the internet, a figure similar to European countries (87%) and close to the developed economies (90%).[11] In addition, for students' lack of access to technological devices, grants and equipment loans were provided in order to foster students' participation in the online courses.

Because COVID-19 broke out right after the start of the academic year, students' decisions about enrollment and registration for courses in the first semester were made before knowing that the classes were going to be online. In addition, the pandemic hit Uruguay more heavily in the second semester of the year; therefore, the virtual modality continued during that period. Moreover, by the end of 2020 and beginning of 2021, COVID-19 cases and deaths were at a peak, leading to the authorities' decision to continue with online courses also in 2021. In contrast with 2020, when enrollment was not affected by COVID-19 and online teaching, in 2021, new students could enroll and register for courses knowing that classes would be online at least for the first semester and with a hybrid modality for the second semester.

6 Data

6.1 Administrative records between 2017-2020

The data comes from different sources. On the one hand, I use administrative records from UDELAR for the period 2017 to 2020.[12] This information includes two different datasets of new enrollments. First, information from the registration form that students fill out when they enroll at the university at the beginning of the year. Com-

[11] The source of the figures is the International Telecommunication Union (ITU) World Telecommunication/ICT Indicators Database

[12] The information was provided by the General Planning Office from UDELAR, and was obtained from the Administrative Management System for Education.

pleting the form is mandatory, and from there, I obtain students' socio-economic and sociodemographic characteristics, such as gender, age, the place where they studied high school, and if the institution was private or public. Second, a dataset with students' records of academic events, i.e., courses enrollment, courses approved, and grades. This information allows me to capture the academic trajectory of students over time. On the other hand, additional socio-economic characteristics such as parental education, students' parenthood, or the number of household members are obtained from a self-administered questionnaire collected yearly. Although completing this form is compulsory, due to the COVID-19 pandemic, the enforcement of this obligation was looser in 2020.

I combine the sources of information detailed before to obtain the final dataset of analysis. From the total population of new enrollments per year, I consider only freshmen in majors with more than 50 students enrolled per year. In addition, my main estimation is restricted to students below 30 years old (85% of the sample) and without any previous enrollment at university (65% of those). This decision is based on the fact that students that were previously enrolled at the university in another degree could potentially already be settled in the place where the campus is located. Because I am interested in analyzing decisions at the student level, if a student enrolled in more than one major, she is only considered in the degree in which she has more courses enrollment.

I use these administrative records to obtain outcomes regarding academic performance. Firstly, I can observe whether students enrolled in the university but did not do any academic activity during the first year of university. I define the variable "No Activity" as a dummy that equals one if the student did not take any final or midterm evaluation during the academic year and zero otherwise. I conceptualized this variable as a measure of dropout. Secondly, I sum the number of courses for which the student took at least one evaluation test by the year, hereafter "Number of courses". As a third outcome, I sum the number of approved subjects during the year, "Number of approved subjects". Finally, I consider the "Mean Grade" as the average of all grades in the transcripts.

In Table 1, I present the main characteristics and the distribution of observations by year for the estimation sample. The first thing to notice is that in the analyzed period, the characteristics of students are stable across the years. 60% of students are women, and the mean age at enrollment is 19 years old. In addition, most university students come from public high school institutions. Considering only the administrative information from students, I have a sample of close to 14.000 students per year. The self-administered questionnaire shows that 80% of the students are white, the vast majority do not have kids when entering university, 20% of the students work, and for 20% of students, at least one of their parents has a university degree. The average household size is 3. The information from the self-questionnaire is useful, but for 2020 and for 2017, there are many missing observations (12% compared with 5% for 2018 and 2019). The non-response to this questionnaire could be associated with less commitment to the university, generating a bias in the sample when considering the self-questionnaire control variables. Because the variables from the questionnaire could be predictors of my outcomes, I estimate the main results in two ways, first, using only the administrative controls and then considering both administrative and self-questionnaire controls.[13]

Table 1: Descriptive statistics of control variables

	2017	2018	2019	2020	Total
Gender(1=Woman)	61.0	60.7	60.1	60.0	60.4
Age at enrollment (degree)	19.5	19.5	19.4	19.4	19.5
Private high-school	23.0	21.7	21.5	19.6	21.4
Public high-school	75.3	76.1	75.9	77.0	76.1
High-school abroad	1.6	2.2	2.6	3.4	2.5
N obs with admin controls	13,892	14,036	14,646	14,650	57,224
Ethincity(1=Non-white)	19.2	19.5	20.8	21.0	20.1
No kids	98.0	97.6	97.5	97.9	97.7
1 Child	1.6	1.9	1.9	1.7	1.8
More than 1 child	0.4	0.5	0.6	0.5	0.5
Work	19.6	20.9	19.5	18.2	19.5
Father or mother with univ.	22.1	21.4	22.0	21.2	21.7
Household size	3.4	3.0	3.0	3.0	3.1
W/o self administred quest.	12.3	5.1	5.6	11.6	8.6
N obs with full controls	11,974	13,088	13,422	12,733	51,217

Notes: The table shows the percentage of students according to the characteristics defined in the first column by year and total, except rows "N obs with admin controls" and "N obs with full controls", which are the number of observations by year and total.

6.2 Enrollment at the locality level for 2021

The data from administrative records give me information regarding the students that decided to enroll at university. However, to analyze the effect of online learning on the decision to attend university, I also need information on those who decided not to enroll. Because I miss this information, I do the analysis at a more aggregate level (locality) measuring changes in enrollment rates that could reflect changes in individual decisions. As mentioned before, localities are geographical units defined in terms of clearly and precisely delimited territories made up of clusters of buildings and therefore reflect the representation of landscape changes. They are characterized by a concentration of population and dwellings.[14]

To compute the enrollment rates, I combine two sources of information. First, the enrollment information comes from the registration form detailed above for the period 2017-2021. I aggregate enrollment by localities and obtain the number of new students enrolled in the university by locality and year. Second, I compute the total number of individuals between 17 and 29 by locality using the Uruguayan Census from 2011, the last one available. I merge both sources of information and compute the share of enrollment on population by locality and year.

6.3 Treatment variable: campus availability

As mentioned before, treatment is given by the fact that for the treated group, COVID-19 and the subsequent switch to online learning, implied the possibility of returning

[13]For the rest of the tables, I only present the results with administrative controls, but the results are similar when adding the self-questionnaire controls

[14]More information about localities is presented in Appendix A

to (or avoiding leaving) their hometowns and/or reducing commuting long distances. Because I do not have information on where they lived in the previous year, I define which students are in the treated or control groups based on where they did the last year of high school. To do this, I recover the high school's locality using the institution's name. Given that the average age of entrance is 19 years old (Table 1), the high school institution should be a good proxy for residence before university.

Based on the previous information, I compute three alternative definitions according to the distance to the campus (*Campus*). First, I consider as treated those students living outside a locality with a campus for their last year of high school (*Outsideloc.*). Second, I consider as treated those students living more than 20 Km from a campus (>20Km). Third, I consider as treated the students living more than 50 Km from a university (>50Km) and as controls the students living less than 20 Km from campus (I do not consider students living between 20 and 50 Km because it is not clear if they are treated or control students). Table 2 shows the distribution of the treated group over time and the total students considered in both the treatment and control groups. As the Table shows, close to 40% of students studied high school in a locality without a university campus, and of those students that did high school living less than 20 Km or more than 50 Km, only 29% are treated.

Table 2: Treatment variable - Student level

	2017	2018	2019	2020	Total
Outside loc.					
Treated	42	43	43	44	43
Total N	13,892	14,036	14,648	14,650	71,154
>20 Km					
Treated	37	38	38	38	38
Total N	13,892	14,036	14,648	14,650	71,154
>50 Km					
Treated	29	29	29	29	29
Total N	12,283	12,359	12,819	12,697	62,245

Notes: The Table shows the percentage of treated (Treated) students and the number of observations in the treated and control groups (Total N) according to different definitions of treatment by year and total. *Outsideloc.* defines treatment considering if students did their high school in a locality without a university campus (treated) and 0 otherwise (controls). >20Km considers as treated students that did high school more than 20 Km from campus and 0 otherwise. >50Km define as treated students those who did high school more than 50 Km from a university campus and controls the students who did high school less than 20 Km from campus.

To analyze enrollment in 2021, I construct the same variables at the locality level. For my main estimation, I used those localities for which at least one student registered at university in the period 2017-2020. Table 3 presents the distribution of treatment at the locality level. Because there are only eight localities with a university campus (the capitals of the eight departments with a university campus), the control group represents only 1% of total localities in the first approach. In addition, I also estimate the regression using all localities with at least 5,000 (the threshold for a place to be considered urban) or 2,000 inhabitants according to the 2011 Census.[15]

[15] In these specifications I ease the restriction of considering localities for which at least one student registered to university. This means that I also use as treatment group places where any student enrolled at the university never for the whole period of analysis.

Table 3: Treatment variable - Locality level

	2017	2018	2019	2020	Total
Outside loc.					
Treated	99	99	99	99	99
Total N	541	541	541	541	2,705
>20 Km					
Treated	87	87	87	87	87
Total N	541	541	541	541	2,705
>50 Km					
Treated	83	83	83	83	83
Total N	421	421	421	421	2,105

Notes: The Table shows the percentage of treated localities (Treated) and the number of observations in the treated and control groups (Total N) according to different definitions of treatment by year and total. *Outsideloc.* defines as treated localities those with a university campus. In the >20Km case, localities with a campus more than 20Km away are treated, and localities with a campus less than 20 Km are controls. >50Km define as treated those localities with a campus more than 50 Km away and controls the localities with a campus less than 20 Km away.

7 Empirical Strategy

To estimate the differential effect of the COVID-19 pandemic and the shift to online learning on students' academic outcomes among the treated and control group, I follow Rodríguez-Planas (2022*a*) framework and estimate the following difference-in-differences model:

$$y_{il} = \beta_0 + \beta_1 Year2020 + \beta_2 (Year2020 * Campus_l) + \gamma_l + \beta_4 X_{il} + \epsilon_{il} \qquad (1)$$

where y_{il} is the outcome of interest for student i in locality l[16]. $Year2020$ is a dummy equal to one for 2020 and 0 before that year. $Campus_l$ is an indicator variable with value one for the treated group, as mentioned before. γ_l represents the locality fixed effects, and X_{il} are the control variables from the registration form (gender, age at enrollment, and type of high school institution) and the self-administered questionnaire (ethnicity, categorical variable for number of kids, it the student has a job, if at least one of the student parents went to the university, and the household size) defined in Section 6. I cluster standard errors at the locality level.

The coefficient of interest, β_2, captures the differential post-pandemic effect on the outcome, y_{il}, for students that are from localities where there is no university campus relative to peers from localities where there is a campus. Because I include locality fixed effects to control for time-invariant observable and unobservable characteristics at that level, the campus indicator is omitted The coefficient β_1 captures the changes in the outcome variables in 2020. Including the control variables allow me to control for observed characteristics of the students.

[16] My data is a pool of repeated cross-section for different years, therefore, I do not include the subindex t in the specification, as each student is observed only once

The empirical strategy of difference-in-differences relies on the identifying assumption of parallel trends across groups. To assess the validity of this assumption, I estimate the following equation using the event study framework to check for preexisting trends:

$$y_{il} = \mu_0 + \sum_{t=2017}^{2020} \mu_t Year_t + \sum_{t=2017}^{2020} \rho_t (Year_t * Campus_l) + \gamma_l + \mu_4 X_{il} + \epsilon_{il} \quad (2)$$

where $Year_t$ is a dummy that takes value one for the year when the outcome was observed and zero otherwise. The $Year_{2019}$ dummy is the reference category. The rest of the variables are defined as before. In the absence of preexisting differential pre-trends, the ρ_t estimated coefficients of years before 2020 should not be statistically different from zero.

To asses if there are differences in the enrollment rate in 2021 I follow a similar strategy, but at the locality level. First, I estimate the following equation:

$$ShEnrollment_{lt} = \theta_0 + \theta_1 Year2021 + \theta_2 (Year2021 * Campus_l) + \gamma_l + \epsilon_{lt} \quad (3)$$

where $ShEnrollment_{lt}$ is the outcome of interest, the share of students enrolled, in locality l and year t. $Year2021$ is a dummy equal to one if the outcome measure is for 2021 and zero before that year. $Campus_l$ is the treatment measure as defined in the previous section, and γ_l represents the locality fixed effects.

In this case, the coefficient of interest, θ_2, captures the differential post-pandemic effect on the outcome, $ShEnrollment_{lt}$, for treated localities relative to control localities. Once again, because I include locality fixed effects to control for time-invariant observable and unobservable characteristics at that level, the campus indicator is omitted.

In addition, I estimate the following equation using the event study framework to assess the validity of the parallel trend assumption:

$$ShEnrollment_{lt} = \sigma_0 + \sum_{t=2017}^{2021} \sigma_t Year_t + \sum_{t=2017}^{2021} \kappa_t (Year_t * Campus_l) + \gamma_l + \epsilon_{lt} \quad (4)$$

where $Year_t$ is a dummy that takes value one for the year when the outcome was observed and zero otherwise. The $Year2020$ dummy is the omitted category. The rest of the variables are defined as before. In the absence of preexisting differential trends, the estimated coefficients of the year previous to 2020 should not be statistically different from zero.

Finally, it is worth mentioning one limitation of this strategy. Because treatment turns on simultaneously with the pandemic, treatment and pandemic are not distinguishable in the regression. If the pandemic differentially affected students from localities with and without a university campus by another channel different from online learning, I am capturing both effects. However, the pandemic did not hit Uruguay in terms of infections until the end of 2020, and by the beginning of 2021, the situation was relatively similar across departments[17]. In addition, most of the decisions

[17] For the distribution of COVID-19 cases in Uruguay: https://guiad-covid.github.io/evolucionP7.html

regarding public policy on the pandemic were taken at the country government level. With respect to the consequences of the economic shock, data constraints makes difficult to measure economic rates at the locality level. However, the distribution at the department level comparing 2020 with 2019 provided by CED (2022) is in line with the parallel trend assumption. Overall, it is plausible to believe that my results are mainly driven by the differential effects of the pandemic on tertiary educational outcomes derived from the online learning opportunity.

8 Results

In this section, I present the results of the analysis. First, I show the results of the effects on the academic outcomes for new first-year students up to 29 years old. I also present a heterogeneous effects analysis. Second, I follow the same order to show the results on the enrollment rates by localities.

8.1 Academic outcomes 2020

In Table 4 I present the main results regarding the effects of COVID-19 and online learning on the academic outcomes of freshmen under 30. In the first place, it is worth mentioning that the pandemic affected all first-year students. According to Panel a, there was an increase of 5.7 pp of enrolled students that did not do any activity, i.e., that dropout from university. Whereas this coefficient shows the effect of COVID-19 on the outcomes, the interaction term reflects the differences due to treatment. Then, the treatment softened the negative impact of the pandemic on the treated group by 1.5 pp compared to the control students. In other words, students from localities far away from the campus, that could potentially return home or commute less, had a lower dropout rate than students from localities where there was already a university campus. Regarding the academic outcomes, conditional on continuing studying, the pandemic had a positive effect on the number of approved subjects (one-third of a course more) and the mean grade (0.6 on a scale that goes from 0 to 12), but there was no differential effect according to treatment. The results hold when including sociodemographic controls obtained from the self-reported questionnaire.

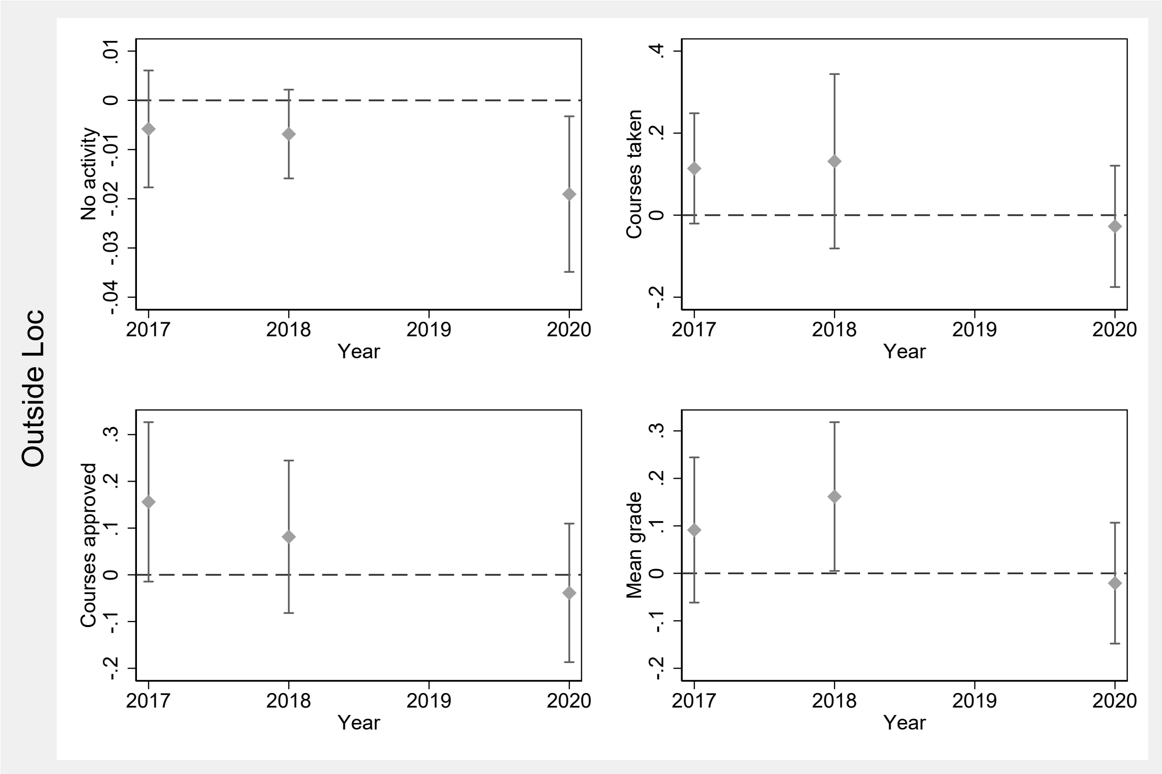
Table 4: Academic outcomes 2020

	No Activity (1)	Number of Courses (2)	Number of Approved subjects (3)	Mean Grade (4)
Panel a: New students up to 29 years old - Administrative controls				
Year2020	0.057***	-0.006	0.348***	0.608***
	(0.007)	(0.075)	(0.057)	(0.046)
	[0.000]	[0.941]	[0.000]	[0.000]
Year2020*Campus_{L}	-0.015*	-0.108	-0.116	-0.103
	(0.008)	(0.084)	(0.071)	(0.067)
	[0.068]	[0.201]	[0.103]	[0.127]
N. Observations	55,342	50,746	50,746	44,153
Panel b: New students up to 29 years old - All controls				
Year2020	0.033***	0.079	0.513***	0.641***
	(0.006)	(0.073)	(0.054)	(0.049)
	[0.000]	[0.281]	[0.000]	[0.000]
Year2020*Campus_{L}	-0.014*	-0.102	-0.076	-0.033
	(0.007)	(0.084)	(0.076)	(0.063)
	[0.053]	[0.223]	[0.320]	[0.595]
N. Observations	49,779	47,137	47,137	41,939

Notes: Reported estimates are obtained from an OLS regression, including locality fixed effects and student control variables. In Panel a, I only include the control variables from the administrative form: gender, age at enrollment, and type of high school institution. In Panel b, I also add the control variables from the self-administered questionnaire: ethnicity, categorical variable for the number of kids, if the student has a job, if at least one of the student's parents went to the university, and the household size. Standard errors reported in parentheses, clustered at the locality level using Liang-Zeger cluster robust standard errors. P-Values are reported in square brackets and obtained using Liang-Zeger cluster robust standard errors. For hypothesis testing we use P-values with significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Columns (1) to (4) refer to the academic outcomes of students obtained from the administrative records as defined in Section 6.Year2020 is a dummy variable that equals 1 for students enrolled in 2020 and 0 otherwise. Year2020*Campus_{L} takes the value 1 for students enrolled in 2020 from the treated group defined as students that did high school in a locality without a university campus. The regression includes new students up to 29 years old.

As I mentioned in the empirical strategy, to show that the difference-in-differences strategy framework can be used in this setting, I present the results of the event study analysis. As figure 2 shows, there were no differences for previous years almost in any of the variables analyzed. Again, the only outcome affected by treatment was the decision of dropout.

Figure 2: Event study analysis for Academic outcomes 2020



Note: These figures plot the coefficients on the interaction between the year2020 and distance to campus treatment (and the 95% confidence intervals) from the regression of the model defined in equation 2. Students that did high school in a locality without a university campus are considered as treated students. Each figure represents the coefficients from the regression on the four different outcome variables considered in the analysis (No activity, Courses taken, Courses approved, and Mean grade). Standard errors clustered at the locality level using Liang-Zeger cluster robust standard errors.

As robustness checks, I also estimate the main equation but just considering degrees with more than 500 students and without degrees with restrictions for entrance or with changes in their curriculum in the period analyzed (Panel a and b of Table B1 respectively). In both cases, results are robust to these specifications.

8.1.1 Heterogeneous effects

First, I estimate the main equation using different distance variables as treatment (Table 5). The results when using as treated students those living more than 20 Km from a university campus are qualitatively similar to those from the main estimation. However, when considering as treated those students more than 50 Km from campus, the coefficient for the interaction doubled my main coefficient. This implies that for those students farthest away from campus, treatment softened dropout rates more pronouncedly.

Table 5: Academic outcomes 2020

	No Activity (1)	Number of Courses (2)	Number of Approved subjects (3)	Mean Grade (4)
Panel a: Treatment: >20Km				
Year2020	0.057***	-0.011	0.341***	0.596***
	(0.007)	(0.069)	(0.053)	(0.043)
	[0.000]	[0.874]	[0.000]	[0.000]
Year2020*Campus$_L$	-0.016**	-0.111	-0.114	-0.085
	(0.008)	(0.083)	(0.071)	(0.069)
	[0.040]	[0.182]	[0.111]	[0.222]
N. Observations	55,342	50,746	50,746	44,153
Panel b: Treatment: >50Km				
Year2020	0.055***	-0.020	0.339***	0.596***
	(0.007)	(0.070)	(0.054)	(0.042)
	[0.000]	[0.776]	[0.000]	[0.000]
Year2020*Campus$_L$	-0.027***	-0.124	-0.141*	-0.108
	(0.008)	(0.092)	(0.079)	(0.073)
	[0.001]	[0.183]	[0.075]	[0.144]
N. Observations	48,274	44,251	44,251	38,574

Notes: Reported estimates are obtained from an OLS regression, including locality fixed effects and student control variables (gender, age at enrollment, and type of high school institution). Standard errors reported in parentheses clustered at the locality level using Liang-Zeger cluster robust standard errors. P-Values are reported in square brackets and obtained using Liang-Zeger cluster robust standard errors. For hypothesis testing I use P-values with significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Columns (1) to (4) refer to the academic outcomes of students obtained from the administrative records as defined in Section 6. Year2020 is a dummy variable that equals 1 for students enrolled in 2020 and 0 otherwise. Year2020*Campus$_L$ takes the value 1 for students enrolled in 2020 from the treated group. In Panel a, the treated group consists of students that did high school more than 20 Km away from a university campus and the control group of other students. In Panel b, the treated group is composed of students living more than 50 Km away from a university campus, and the control group of students living less than 20 Km from a university campus. The regression includes new students up to 29 years old.

I also estimate the equation using different subsamples according to age and the previous institutional link with the university. Panel a from Table 6 shows the results for those new students up to 25 years old. The results are very similar to the results for the whole sample. On the opposite side, results differ when I consider the whole sample of first-year students enrolled in the university instead of only new students (as in my main estimation). First-year sample includes those students who enrolled in a career for the first time but could already have been enrolled in another career before. The results considering all students (Panel b and c) show that there is no differential effect of treatment in the dropout decisions, and I observe relatively slightly worst results for the other variables. This could be related to the fact that the treatment variable reflects the place where students did high school and not the previous residence. Including students previously enrolled at the university (as I do in Panel b and c) could mean that I am considering students already settled in a place with a university campus as treated.

Table 6: Academic outcomes 2020

	No Activity (1)	Number of Courses (2)	Number of Approved subjects (3)	Mean Grade (4)
Panel a: New students up to 25 years old				
Year2020	0.058***	-0.023	0.333***	0.582***
	(0.007)	(0.074)	(0.058)	(0.050)
	[0.000]	[0.754]	[0.000]	[0.000]
Year2020*Campus$_L$	-0.016*	-0.080	-0.094	-0.093
	(0.008)	(0.082)	(0.069)	(0.067)
	[0.054]	[0.329]	[0.175]	[0.171]
N. Observations	52,230	48,182	48,182	42,268
Panel b: All students up to 29 years old				
Year2020	0.051***	-0.480***	0.260***	0.873***
	(0.004)	(0.064)	(0.047)	(0.050)
	[0.000]	[0.000]	[0.000]	[0.000]
Year2020*Campus$_L$	-0.002	-0.150**	-0.146**	-0.161***
	(0.006)	(0.076)	(0.062)	(0.060)
	[0.769]	[0.049]	[0.020]	[0.008]
N. Observations	82,125	72,738	72,738	61,740
Panel c: All students up to 25 years old				
Year2020	0.053***	-0.462***	0.331***	0.919***
	(0.004)	(0.069)	(0.045)	(0.055)
	[0.000]	[0.000]	[0.000]	[0.000]
Year2020*Campus$_L$	-0.006	-0.113	-0.128**	-0.153**
	(0.007)	(0.077)	(0.060)	(0.062)
	[0.406]	[0.145]	[0.036]	[0.014]
N. Observations	73,437	66,055	66,055	56,876

Notes: Reported estimates are obtained from an OLS regression, including locality fixed effects and student control variables (gender, age at enrollment, and type of high school institution). Standard errors reported in parentheses clustered at the locality level using Liang-Zeger cluster robust standard errors. P-Values are reported in square brackets and obtained using Liang-Zeger cluster robust standard errors. For hypothesis testing I use P-values with significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Columns (1) to (4) refer to the academic outcomes of students obtained from the administrative records as defined in Section 6. Year2020 is a dummy variable that equals 1 for students enrolled in 2020 and 0 otherwise. Year2020*Campus$_L$ takes the value 1 for students enrolled in 2020 from the treated group defined as students that did high school in a locality without a university campus. In Panel a, the regression only includes new students up to 25 years old. In Panel b, the regression includes all students up to 29 years old. In Panel c, the regression includes all students up to 25 years old.

Previous literature analyzing the effects of the pandemic shows that there could be differences according to gender and socioeconomic background of students. Therefore, I run the main equation separately for boys and girls to capture differences by gender. Table B2 shows that when considering those students living outside a locality with a campus, dropout results are qualitatively similar to the main estimation in terms of the coefficient magnitude but only significant for girls. However, when I consider being more than 20 Km away from campus as treatment the effect, the effect is similar for girls and becomes significant and more pronounced for boys (Table B3). The more pronounced positive effect for girls goes in line with some of the papers studying the impact of online learning during COVID-19 by gender (Aucejo et al. (2020), Kofoed et al. (2021)). These papers find that girls prefer online learning more than boys.

However, my results when the treatment variable considered being more than 20 Km away could suggest that distance may become online learning a solution for everyone.

To measure the socioeconomic background, I use the type of high school institution (private or public) where the student did secondary education. I observe that for both treatment variables, the differential effect of treatment in the decrease of dropout is driven mainly by students from public high schools (Table B4 and Table B5). This could suggest that students from less affluent socioeconomic backgrounds respond more to a reduction in costs associated with distance to a university campus. However, it is worth mentioning that enrollment in private high schools is particularly low in the treated group (6% of treated students).

8.2 Enrollment in 2021

As I mentioned, in 2021, both the enrollment and courses were online since the beginning of the year, and this decision was communicated and disseminated institutionally. The possibility of enrolling and attending classes virtually could have led to an increase in enrollment in places far away from campuses. In this section, I present the results of that analysis.

Table 7 shows the estimated results of equation 3, where I measure the differential effect of the pandemic and online learning on the share of enrollment of new students up to 29 years old by locality. Results show an increase of 0.4 pp in the share of enrollment in localities without a university campus. This effect represents an increase of 13% in the share of enrollment of the treated localities. When analyzing the results according to the distance to campus, I observe that the effect is stable for all treatment variables. Results suggest that online learning leads to an increase in university enrollment, thus building on the idea that distance matters in students' enrollment decisions.

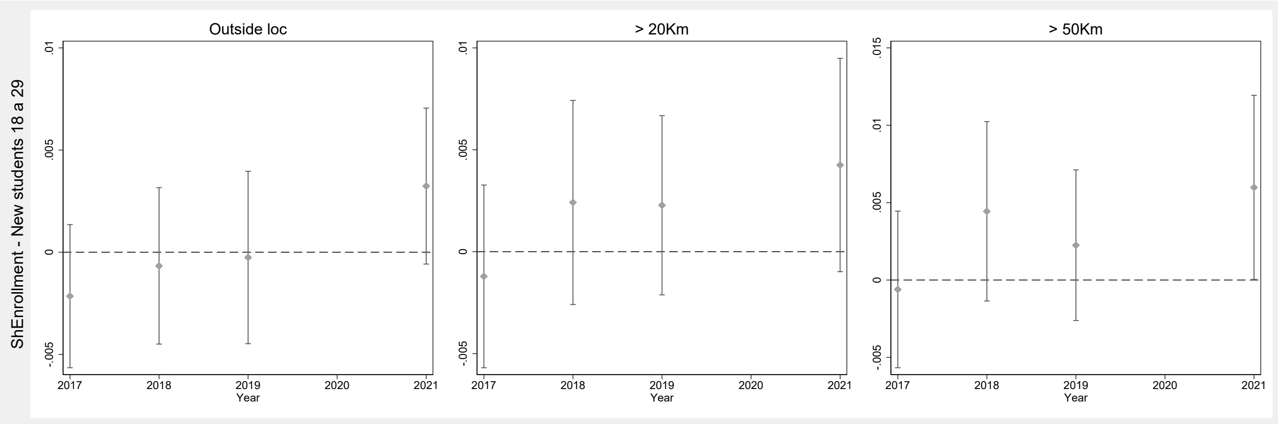
Table 7: Enrollment 2021 - New students up to 29 years old

	Share of enrollment		
	Outside loc	> 20Km	> 50Km
	(1)	(2)	(3)
	Outside loc	> 20Km	> 50Km
Year2021	-0.001	0.000	0.000
	(0.001)	(0.002)	(0.002)
	[0.107]	[0.930]	[0.930]
Year2021*Campus$_L$	0.004**	0.003	0.004*
	(0.002)	(0.002)	(0.003)
	[0.012]	[0.154]	[0.092]
N. Observations	645	645	490

Notes: Reported estimates are obtained from an OLS regression, including locality fixed effects. Standard errors reported in parentheses clustered at the locality level using Liang-Zeger cluster robust standard errors. P-Values are reported in square brackets and obtained using Liang-Zeger cluster robust standard errors. For hypothesis testing I use P-values with significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The outcome variable is the share of enrollment of students by locality defined as explained in Section 6. The share of enrollment is computed using all students up to 29 years old. Columns (1) to (3) differ in how the treatment groups are composed. In (1), the treated group consists of students that did high school in a locality without a university campus. In (2), the treated students are students that did high school more than 20 Km away from a university campus, and the control group by the other students. In (3), the treated group is composed of students living more than 50 Km away from a university campus, and the control group of students living less than 20 Km from a university campus. Year2021 is a dummy variable that equals 1 for localities in 2021 and 0 otherwise. Year2021*Campus$_L$ takes the value of one for localities in 2021 from the treated group defined as explained before. The regression includes all localities for which at least one student enrolled in 2017-2021.

Figure 3 shows the event study analysis to provide evidence in favor of the parallel trends assumption.

Figure 3: Event study analysis for Enrollment 2021



Notes: These figures plot the coefficients on the interaction between the year2021 and distance to campus treatment (and the 95% confidence intervals) from the regression of the model defined in equation 4 where the outcome variable is the share of enrollment. Each figure considers a different definition of distance to campus. The first one considered as treated, localities where the university campus is outside the locality. The one in the middle considers as treated localities more than 20 Km away from a university campus. The third figure considers as treatment the localities more than 50Km away from a university campus and as control localities less than 20 Km away from a university campus. Standard errors clustered at the locality level using Liang-Zeger cluster robust standard errors.

8.2.1 Heterogeneous effects

As before, I analyze if there are differences according to age or previous link with the university. To do this, I compute the enrollment rates by localities considering enrollment of the different analyzed groups. Panel a of Table 8 shows that if I consider all students up to 29 years old and not only those without previous enrollment in the university, the effects are more pronounced and significant for all the treatment definitions. This could be capturing the fact that the switch to online courses widens the degree offer also for students with a previous linkage with UDELAR. On the other side, Panel b shows that the effect is similar for new students up to 25 than those up to 29.

Table 8: Enrollment 2021

	Campus distance		
	Outside loc	> 20Km	> 50Km
	(1)	(2)	(3)
Panel a: All students up to 29 years old			
	Outside loc	> 20Km	> 50Km
Year2021	0.002*	0.003*	0.003*
	(0.001)	(0.002)	(0.002)
	[0.064]	[0.082]	[0.083]
Year2021*Campus_{L}	0.005***	0.005**	0.006**
	(0.002)	(0.002)	(0.003)
	[0.007]	[0.046]	[0.047]
N. Observations	645	645	490
Panel b: New students up to 25 years old			
	Outside loc	> 20Km	> 50Km
Year2021	-0.001*	0.000	0.000
	(0.001)	(0.002)	(0.002)
	[0.082]	[0.839]	[0.839]
Year2021*Campus_{L}	0.005**	0.003	0.005
	(0.002)	(0.003)	(0.003)
	[0.021]	[0.260]	[0.146]
N. Observations	645	645	490

Notes: Reported estimates are obtained from an OLS regression, including locality fixed effects. Standard errors reported in parentheses clustered at the locality level using Liang-Zeger cluster robust standard errors. P-Values are reported in square brackets and obtained using Liang-Zeger cluster robust standard errors. For hypothesis testing I use P-values with significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The outcome variable is the share of enrollment of students by locality defined as explained in Section 6. In Panel a, the share of enrollment is computed using all students up to 29 years old. In Panel b, the share of enrollment is computed using new students up to 25 years old. Columns (1) to (3) differ in how the treatment groups are composed. In (1), the treated group consists of students that did high school in a locality without a university campus. In (2), the treated students are students that did high school more than 20 Km away from a university campus, and the control group by the other students. In (3), the treated group is composed of students living more than 50 Km away from a university campus, and the control group of students living less than 20 Km from a university campus. Year2021 is a dummy variable that equals 1 for localities in 2021 and 0 otherwise. Year2021*Campus_{L} takes the value of one localities in 2021 from the treated group defined as explained before. The regression includes all localities for which there is at least one student enrolled in 2017-2021.

In addition, I estimate the main results considering all urban localities (Table B6) and localities of more than 2000 inhabitants (Table B7). In these specifications, and differently than in my main estimation, I also include localities without any students enrolled in the 2017-2021 period. This implies that in the treated group, I include localities with enrolment rates equal to zero for the whole period. When including these changes, the results remain stable when considering treatment as having a university campus in the locality. However, I do not observe any effect when accounting for distance in the treatment. This could suggest the relevance of knowing older university students when deciding to continue studying. In line with this idea, Pistolesi (2022) stress the importance of peer effects in the decision of enrollment to university, and

Bobonis & Finan (2009) for secondary school enrollment.

9 Final Remarks

Using administrative data from a public university in Uruguay, I analyze the differential effect of COVID-19 and the consequent online learning shift on freshman academic outcomes students according to the distance from the university. This setting allows me to contribute to understanding the potential benefits of online learning in reducing the distance (or commuting time) to university.

I find that the pandemic increased dropouts, but students that now could avoid being far away from home had a lower dropout rate. This effect holds when using different measure of distances as the treatment variable. In addition, conditional on continuing the university, there is no systematic effect on other academic outcomes (such as the number of courses, approved subjects or mean grade). In addition, I analyze the effect on the decision to attend university. To do this, I aggregate the information at the locality level to compute enrollment rates by locality and year. I find that there was an increase in the enrollment rate in those localities without a university campus. Again, this stress the importance of distances in the decision of university enrollment and online learning as a potential solution.

These findings shed light on a possible answer to reducing geographical inequalities in access to tertiary education. This is particularly relevant for the developing world, where tertiary education rates are lower. However, it is worth mentioning that connectivity throughout the territory is needed to take advantage of online learning. Uruguay constitutes an interesting case to study because it is a developing country with tertiary enrollment rates that are still below developed economies but with a high internet connectivity figure. The economy of scale of providing access to tertiary education via online learning for those far away from a university campus is a feasible requirement. Therefore, digital inclusion efforts could also increase tertiary education enrollment rates. However, because literature also has shown adverse effects of online learning compared to live teaching (Figlio et al. 2013, Kofoed et al. 2021, De Paola et al. 2022, Bird et al. 2022) and Bettinger et al. (2017), placing online learning as a substitute for in-person classes could also have disadvantages. Overall, there is space to continue contributing to the design of policies to take advantage of new technologies and tackle their drawbacks in the educational system.

References

Alm, J. & Winters, J. V. (2009), 'Distance and intrastate college student migration', *Economics of Education Review* **28**(6), 728–738.
URL: *https://ideas.repec.org/a/eee/ecoedu/v28y2009i6p728-738.html*

Altindag, D. T., Filiz, E. S. & Tekin, E. (2021), Is Online Education Working?, NBER Working Papers 29113, National Bureau of Economic Research, Inc.
URL: *https://ideas.repec.org/p/nbr/nberwo/29113.html*

Aucejo, E. M., French, J., Ugalde Araya, M. P. & Zafar, B. (2020), 'The impact of COVID-19 on student experiences and expectations: Evidence from a survey', *Journal of Public Economics* **191**(C).
URL: *https://ideas.repec.org/a/eee/pubeco/v191y2020ics0047272720301353.html*

Bacher-Hicks, A., Goodman, J. & Mulhern, C. (2021), 'Inequality in household adaptation to schooling shocks: Covid-induced online learning engagement in real time', *Journal of Public Economics* **193**(C).
URL: *https://ideas.repec.org/a/eee/pubeco/v193y2021ics0047272720302097.html*

Bettinger, E. P., Fox, L., Loeb, S. & Taylor, E. S. (2017), 'Virtual Classrooms: How Online College Courses Affect Student Success', *American Economic Review* **107**(9), 2855–2875.
URL: *https://ideas.repec.org/a/aea/aecrev/v107y2017i9p2855-75.html*

Bird, K. A., Castleman, B. L. & Lohner, G. (2022), 'Negative impacts from the shift to online learning during the covid-19 crisis: Evidence from a statewide community college system', *AERA Open* **8**, 23328584221081220.
URL: *https://doi.org/10.1177/23328584221081220*

Bobonis, G. J. & Finan, F. (2009), 'Neighborhood Peer Effects in Secondary School Enrollment Decisions', *The Review of Economics and Statistics* **91**(4), 695–716.
URL: *https://ideas.repec.org/a/tpr/restat/v91y2009i4p695-716.html*

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

Bulman, G. & Fairlie, R. W. (2022), The Impact of COVID-19 on Community College Enrollment and Student Success: Evidence from California Administrative Data, IZA Discussion Papers 15424, Institute of Labor Economics (IZA).
URL: *https://ideas.repec.org/p/iza/izadps/dp15424.html*

Caño-Guiral, M. (2018), World Bank Education Overview : Higher Education (English), Technical report.

CED (2022), Análisis del Mercado de Trabajo a partir de microdatos de la ECH, Technical report.

Chetty, R., Friedman, J. N., Hendren, N., Stepner, M. & Team, T. O. I. (2020), The economic impacts of covid-19: Evidence from a new public database built using private sector data, NBER Working Papers 27431, National Bureau of Economic

Research, Inc.
URL: *https://EconPapers.repec.org/RePEc:nbr:nberwo:27431*

De Paola, M., Gioia, F. & Scoppa, V. (2022), Online Teaching, Procrastination and Students' Achievement: Evidence from COVID-19 Induced Remote Learning, IZA Discussion Papers 15031, Institute of Labor Economics (IZA).
URL: *https://ideas.repec.org/p/iza/izadps/dp15031.html*

Failache, E., Fiori, N., Katzkowicz, N., Machado, A. & Méndez, L. (2022), Impact of COVID-19 on higher education: Evidence from Uruguay, Documento de Trabajo 02/2022, Instituto de Economía.

Figlio, D., Rush, M. & Yin, L. (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.
URL: *https://ideas.repec.org/a/ucp/jlabec/doi10.1086-669930.html*

Frenette, M. (2009), 'Do universities benefit local youth? Evidence from the creation of new universities', *Economics of Education Review* **28**(3), 318–328.
URL: *https://ideas.repec.org/a/eee/ecoedu/v28y2009i3p318-328.html*

Gibbons, S. & Vignoles, A. (2012), 'Geography, choice and participation in higher education in England', *Regional Science and Urban Economics* **42**(1-2), 98–113.
URL: *https://ideas.repec.org/a/eee/regeco/v42y2012i1p98-113.html*

Hossain, M. (2021), 'Unequal experience of COVID-induced remote schooling in four developing countries', *International Journal of Educational Development* **85**(C).
URL: *https://ideas.repec.org/a/eee/injoed/v85y2021ics0738059321000997.html*

Jaeger, D. A., Arellano-Bover, J., Karbownik, K., Martínez Matute, M., Nunley, J. M., Seals Jr., R. A., Almunia, M., Alston, M., Becker, S. O. & Beneito, P. (2021), The Global COVID-19 Student Survey: First Wave Results, IZA Discussion Papers 14419, Institute of Labor Economics (IZA).
URL: *https://ideas.repec.org/p/iza/izadps/dp14419.html*

Jardim, G. (2020), How the Availability of Higher Education Affects Incentives? Evidence from Federal University Openings in Brazil, Papers 2011.03120, arXiv.org.
URL: *https://ideas.repec.org/p/arx/papers/2011.03120.html*

Katzkowicz, N., Querejeta, M. & Rosá, T. (2021), Spatial inequalities in educational opportunities: The role of public policies, Technical report.

Kofoed, M. S., Gebhart, L., Gilmore, D. & Moschitto, R. (2021), Zooming to Class?: Experimental Evidence on College Students' Online Learning during COVID-19, IZA Discussion Papers 14356, Institute of Labor Economics (IZA).
URL: *https://ideas.repec.org/p/iza/izadps/dp14356.html*

Lapid, P. (2016), Expanding college access: The impact of new universities on local enrollment, Technical report, Job Market Paper.

Pistolesi, N. (2022), 'Enrolling at university and the social influence of peers', *IZA Journal of Labor Economics* **11**(1), –.
URL: *https://doi.org/10.2478/izajole-2022-0001*

Rodríguez-Planas, N. (2022*a*), 'COVID-19, college academic performance, and the flexible grading policy: A longitudinal analysis', *Journal of Public Economics* **207**(C).
URL: *https://ideas.repec.org/a/eee/pubeco/v207y2022ics0047272722000081.html*

Rodríguez-Planas, N. (2022*b*), 'Hitting where it hurts most: COVID-19 and low-income urban college students', *Economics of Education Review* **87**(C).
URL: *https://ideas.repec.org/a/eee/ecoedu/v87y2022ics0272775722000103.html*

Spiess, C. K. & Wrohlich, K. (2010), 'Does distance determine who attends a university in Germany?', *Economics of Education Review* **29**(3), 470–479.
URL: *https://ideas.repec.org/a/eee/ecoedu/v29y2010i3p470-479.html*

Udelar (2020), Propuesta de la Udelar al país 2020-2024: Plan estratégico de desarrollo, Technical report.

10 Appendix

A Geographic details of Uruguay

A.1 Information at the department level

Uruguay's surface is 176.215 Km^2, and it has different geographical divisions. The more aggregate geographical divisions are departments, of which there are 19. Figure A.1 shows the division of Uruguay according to the departments and the total population per department. Figure A.3 shows the density by department.

Figure A.1: Total population by department

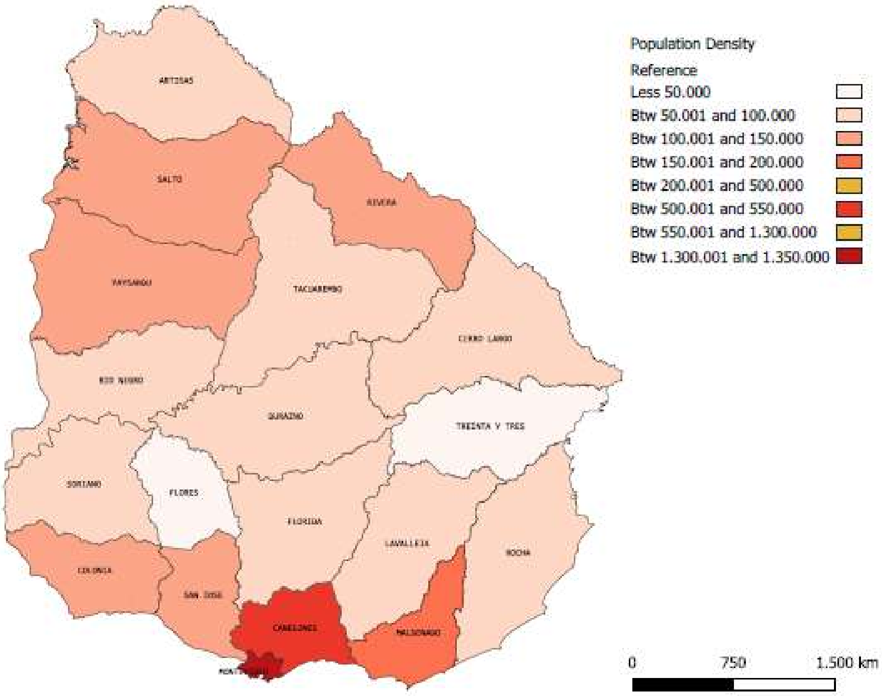
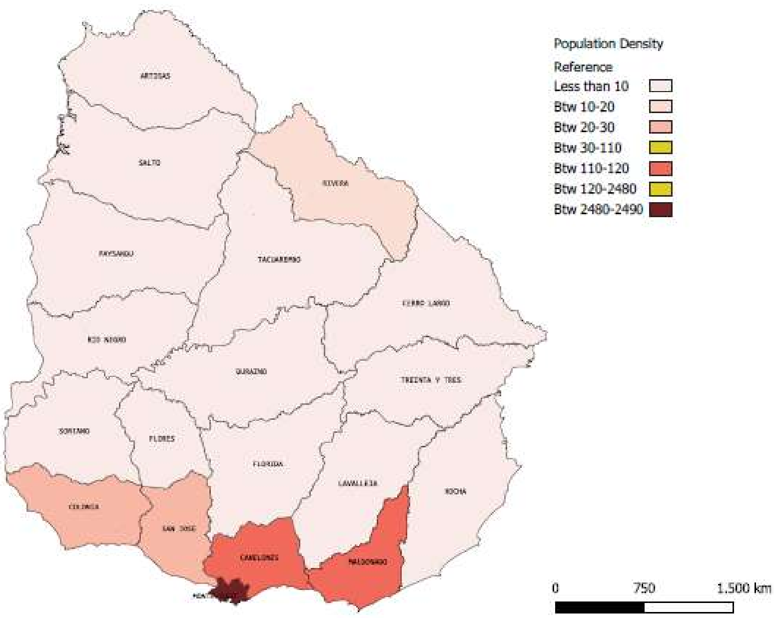


Figure A.2: Density by department

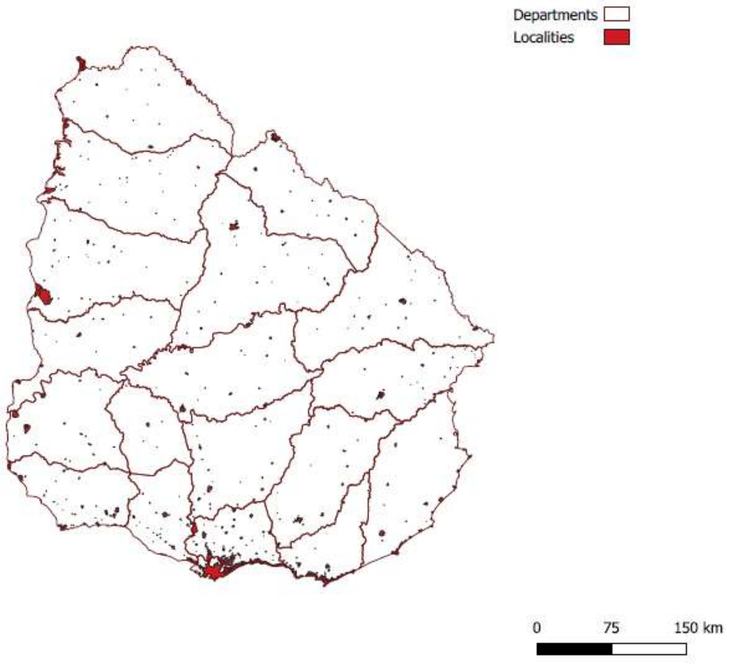


B Information at the locality level

According to the National Institute of Statistics, geographical localities (or census localities) are defined in terms of clearly and precisely delimited territories made up of clusters of buildings and therefore reflect the representation of landscape changes. A Census Locality corresponds to a set of census tracts characterized by a concentration of population and dwellings.

Uruguay has 615 localities based on the information of 2011 Census from the National Institute of Statistics. These localities have a median of 296 inhabitants with a high level of dispersion. The mean is 5057 inhabitants. The mean of department localities is 32, with Montevideo in the lower tail (only one locality for the whole department) and Canelones in the upper tail (with 117 localities).

Figure A.3: Localities geographical distribution



C Other estimations for academic outcomes 2020

Table B1: New students up to 29 years old

	No Activity (1)	Number of Courses (2)	Number of Approved subjects (3)	Mean Grade (4)
Panel a: Only degrees of more than 500 students				
Year2020	0.067***	0.039	0.538***	0.698***
	(0.007)	(0.109)	(0.066)	(0.053)
	[0.000]	[0.717]	[0.000]	[0.000]
Year2020*Campus$_L$	-0.018*	-0.159	-0.201**	-0.120*
	(0.010)	(0.148)	(0.086)	(0.061)
	[0.067]	[0.287]	[0.020]	[0.051]
N. Observations	39,161	35,990	35,990	32,060
Panel b: Only degrees without lottery for entrance neither changes in their curriculum				
Year2020	0.072***	-0.076	0.336***	0.716***
	(0.006)	(0.073)	(0.061)	(0.049)
	[0.000]	[0.302]	[0.000]	[0.000]
Year2020*Campus$_L$	-0.015*	-0.047	-0.039	-0.068
	(0.008)	(0.094)	(0.064)	(0.068)
	[0.062]	[0.618]	[0.539]	[0.320]
N. Observations	47,013	42,898	42,898	37,233

Notes: Reported estimates are obtained from an OLS regression including locality fixed effects and student control variables (gender, age at enrollment and type of high school institution). In Panel a, I only consider students registered in degrees of more than 500 students. In Panel b, I only consider students registered in degrees without a lottery for entrance and changes in the curriculum for 2017-2021. Standard errors reported in parentheses, clustered at the locality level using Liang-Zeger cluster robust standard errors. P-Values are reported in square brackets and obtained using Liang-Zeger cluster robust standard errors. For hypothesis testing I use P-values with significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Columns (1) to (4) refer to the academic outcomes of students obtained from the administrative records as defined in section 6. Year2020 is a dummy variable that equals 1 for students enrolled in 2020 and 0 otherwise. Year2020*Campus$_L$ takes the value 1 for students enrolled in 2020 from the treated group defined as students that did high school in a locality without a university campus. The regression includes new students up to 29 years old.

Table B2: New students up to 29 years old by gender - Treatment: Outsise loc.

	No Activity (1)	Number of Courses (2)	Number of Approved subjects (3)	Mean Grade (4)
Panel a: Boys				
Year2020	0.076***	-0.061	0.256**	0.503***
	(0.007)	(0.074)	(0.099)	(0.052)
	[0.000]	[0.408]	[0.011]	[0.000]
Year2020*Campus$_L$	-0.014	-0.182*	-0.132	-0.119
	(0.012)	(0.101)	(0.131)	(0.107)
	[0.258]	[0.074]	[0.314]	[0.265]
N. Observations	21,884	19,971	19,971	16,464
Panel b: Girls				
Year2020	0.044***	0.032	0.420***	0.683***
	(0.007)	(0.104)	(0.072)	(0.050)
	[0.000]	[0.759]	[0.000]	[0.000]
Year2020*Campus$_L$	-0.013*	-0.075	-0.119*	-0.102
	(0.007)	(0.101)	(0.067)	(0.068)
	[0.082]	[0.456]	[0.079]	[0.139]
N. Observations	33,458	30,775	30,775	27,689

Notes: Reported estimates are obtained from an OLS regression including locality fixed effects and student control variables (gender, age at enrollment and type of high school institution). In Panel a, I run the regression only for boys, while in Panel b, I only include girls. Standard errors reported in parentheses, clustered at the locality level using Liang-Zeger cluster robust standard errors. P-Values are reported in square brackets and obtained using Liang-Zeger cluster robust standard errors. For hypothesis testing I use P-values with significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Columns (1) to (4) refer to the academic outcomes of students obtained from the administrative records as defined in section 6. Year2020 is a dummy variable that equals 1 for students enrolled in 2020 and 0 otherwise. Year2020*Campus$_L$ takes the value 1 for students enrolled in 2020 from the treated group defined as students that did high school in a locality without a university campus. The regression includes new students up to 29 years old.

Table B3: New students up to 29 years old by gender - Treatment: >20Km

	No Activity (1)	Number of Courses (2)	Number of Approved subjects (3)	Mean Grade (4)
Panel a: Boys - Treatment: >20Km				
Year2020	0.078***	-0.088	0.240**	0.486***
	(0.007)	(0.084)	(0.097)	(0.052)
	[0.000]	[0.293]	[0.015]	[0.000]
Year2020*Campus$_L$	-0.020*	-0.132	-0.107	-0.087
	(0.012)	(0.112)	(0.132)	(0.112)
	[0.084]	[0.241]	[0.419]	[0.441]
N. Observations	21,884	19,971	19,971	16,464
Panel b: Girls - Treatment: >20Km				
Year2020	0.043***	0.039	0.414***	0.672***
	(0.008)	(0.098)	(0.069)	(0.048)
	[0.000]	[0.695]	[0.000]	[0.000]
Year2020*Campus$_L$	-0.012	-0.106	-0.125*	-0.090
	(0.008)	(0.098)	(0.069)	(0.071)
	[0.130]	[0.282]	[0.071]	[0.208]
N. Observations	33,458	30,775	30,775	27,689

Notes: Reported estimates are obtained from an OLS regression including locality fixed effects and student control variables (gender, age at enrollment and type of high school institution). In Panel a, I run the regression only for boys, while in Panel b, I only include girls. Standard errors reported in parentheses, clustered at the locality level using Liang-Zeger cluster robust standard errors. P-Values are reported in square brackets and obtained using Liang-Zeger cluster robust standard errors. For hypothesis testing I use P-values with significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Columns (1) to (4) refer to the academic outcomes of students obtained from the administrative records as defined in section 6. Year2020 is a dummy variable that equals 1 for students enrolled in 2020 and 0 otherwise. Year2020*Campus$_L$ takes the value 1 for students enrolled in 2020 from the treated group defined as students that did high school in a locality more than 20Km away from a university campus. The regression includes new students up to 29 years old.

Table B4: New students up to 29 years old by socioeconomic background - Treatment: Outsise loc.

	No Activity (1)	Number of Courses (2)	Number of Approved subjects (3)	Mean Grade (4)
Panel a: Public high school				
Year2020	0.062***	-0.115	0.296***	0.609***
	(0.010)	(0.104)	(0.075)	(0.060)
	[0.000]	[0.271]	[0.000]	[0.000]
Year2020*Campus$_L$	-0.017*	0.031	-0.030	-0.065
	(0.010)	(0.100)	(0.076)	(0.071)
	[0.097]	[0.757]	[0.692]	[0.358]
N. Observations	43,236	39,530	39,530	34,187
Panel b: Private high school				
Year2020	0.046***	0.176***	0.397***	0.532***
	(0.003)	(0.055)	(0.051)	(0.040)
	[0.000]	[0.004]	[0.000]	[0.000]
Year2020*Campus$_L$	-0.013	-0.097	-0.066	-0.008
	(0.024)	(0.141)	(0.257)	(0.302)
	[0.574]	[0.496]	[0.801]	[0.978]
N. Observations	12,106	11,216	11,216	9,966

Notes: Reported estimates are obtained from an OLS regression including locality fixed effects and student control variables (gender, age at enrollment and type of high school institution). In Panel a, I run the regression only for students that attended a public high school, while in Panel b, I only include students that attended a private high school. Standard errors reported in parentheses, clustered at the locality level using Liang-Zeger cluster robust standard errors. P-Values are reported in square brackets and obtained using Liang-Zeger cluster robust standard errors. For hypothesis testing I use P-values with significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Columns (1) to (4) refer to the academic outcomes of students obtained from the administrative records as defined in section 6. Year2020 is a dummy variable that equals 1 for students enrolled in 2020 and 0 otherwise. Year2020*Campus$_L$ takes the value 1 for students enrolled in 2020 from the treated group defined as students that did high school in a locality without a university campus. The regression includes new students up to 29 years old.

Table B5: New students up to 29 years old by socioeconomic background - Treatment: Outsise >20Km

	No Activity (1)	Number of Courses (2)	Number of Approved subjects (3)	Mean Grade (4)
Panel a: Public high school - Treatment: >20Km				
Year2020	0.061***	-0.114	0.282***	0.585***
	(0.010)	(0.095)	(0.067)	(0.059)
	[0.000]	[0.229]	[0.000]	[0.000]
Year2020*Campus_L	-0.019**	0.034	-0.004	-0.020
	(0.010)	(0.094)	(0.076)	(0.075)
	[0.050]	[0.720]	[0.959]	[0.789]
N. Observations	43,236	39,530	39,530	34,187
Panel b: Private high school - Treatment: >20Km				
Year2020	0.045***	0.191***	0.428***	0.559***
	(0.004)	(0.046)	(0.044)	(0.054)
	[0.000]	[0.000]	[0.000]	[0.000]
Year2020*Campus_L	-0.007	-0.291***	-0.415*	-0.286
	(0.030)	(0.093)	(0.234)	(0.272)
	[0.821]	[0.005]	[0.090]	[0.305]
N. Observations	12,106	11,216	11,216	9,966

Notes: Reported estimates are obtained from an OLS regression including locality fixed effects and student control variables (gender, age at enrollment and type of high school institution). In Panel a, I run the regression only for students that attended a public high school, while in Panel b, I only include students that attended a private high school. Standard errors reported in parentheses, clustered at the locality level using Liang-Zeger cluster robust standard errors. P-Values are reported in square brackets and obtained using Liang-Zeger cluster robust standard errors. For hypothesis testing I use P-values with significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Columns (1) to (4) refer to the academic outcomes of students obtained from the administrative records as defined in section 6. Year2020 is a dummy variable that equals 1 for students enrolled in 2020 and 0 otherwise. Year2020*Campus_L takes the value 1 for students enrolled in 2020 from the treated group defined as students that did high school in a locality more than 20Km away from a university campus. The regression includes new students up to 29 years old.

D Other estimations for enrollment in 2021

Table B6: Enrollment 2021 for urban localities

	Campus distance		
	Outside loc	> 20Km	> 50Km
	(1)	(2)	(3)
Panel a: All students up to 29 years old			
	Outside loc	> 20Km	> 50Km
Year2021	-0.001	0.001*	0.001*
	(0.001)	(0.001)	(0.001)
	[0.111]	[0.091]	[0.093]
Year2021*Campus$_L$	0.002**	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
	[0.017]	[0.634]	[0.728]
N. Observations	335	335	255
Panel b: New students up to 25 years old			
	Outside loc	> 20Km	> 50Km
Year2021	-0.001*	0.002*	0.002*
	(0.001)	(0.001)	(0.001)
	[0.086]	[0.061]	[0.063]
Year2021*Campus$_L$	0.003**	-0.001	-0.001
	(0.001)	(0.002)	(0.002)
	[0.021]	[0.465]	[0.580]
N. Observations	335	335	255

Notes: Reported estimates are obtained from an OLS regression including locality fixed effects. In Panel a, the outcome variable is the share of enrollment of new students up to 29 years of age over the total population from 17 to 29 years of age by locality. In Panel b, the outcome variable is the share of enrollment of new students up to 25 years of age over the total population from 17 to 29 years of age by locality. Standard errors reported in parentheses clustered at the locality level using Liang-Zeger cluster robust standard errors. P-Values are reported in square brackets and obtained using Liang-Zeger cluster robust standard errors. For hypothesis testing I use P-values with significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Columns (1) to (3) differ in how the treatment groups is composed. In (1), the treated group consists of students that did high school in a locality without a university campus. In (2), the treated students are students that did high school more than 20 Km away from a university campus, and the control group by the other students. In (3), the treated group is composed of students living more than 50 Km away from a university campus, and the control group of students living less than 20 Km from a university campus. Year2021 is a dummy variable that equals 1 for localities in 2021 and 0 otherwise. Year2021*Campus$_L$ takes the value of one for localities in 2021 from the treated group defined as explained before. The regression includes all localities with more than 5,000 inhabitants (urban localities) in 2017-2021.

Table B7: Enrollment 2021 for localities with more than 2000 inhabitants

	Campus distance		
	Outside loc	> 20Km	> 50Km
	(1)	(2)	(3)
Panel a: New students up to 29 years old			
	Outside loc	> 20Km	> 50Km
Year2021	-0.001	0.001*	0.001*
	(0.001)	(0.001)	(0.001)
	[0.108]	[0.088]	[0.090]
Year2021*$Campus_L$	0.002**	-0.000	-0.001
	(0.001)	(0.001)	(0.001)
	[0.020]	[0.704]	[0.693]
N. Observations	500	500	380
Panel b: New students up to 25 years old			
	Outside loc	> 20Km	> 50Km
Year2021	-0.001*	0.002*	0.002*
	(0.001)	(0.001)	(0.001)
	[0.083]	[0.059]	[0.060]
Year2021*$Campus_L$	0.003**	-0.001	-0.001
	(0.001)	(0.002)	(0.002)
	[0.027]	[0.525]	[0.598]
N. Observations	500	500	380

Notes: Reported estimates are obtained from an OLS regression including locality fixed effects. In Panel a, the outcome variable is the share of enrollment of new students up to 29 years of age over the total population from 17 to 29 years of age by locality. In Panel b, the outcome variable is the share of enrollment of new students up to 25 years of age over the total population from 17 to 29 years of age by locality. Standard errors reported in parentheses clustered at the locality level using Liang-Zeger cluster robust standard errors. P-Values are reported in square brackets and obtained using Liang-Zeger cluster robust standard errors. For hypothesis testing I use P-values with significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Columns (1) to (3) differ in how the treatment group is composed. In (1), the treated group consists of students that did high school in a locality without a university campus. In (2), the treated students are students that did high school more than 20 Km away from a university campus, and the control group by the other students. In (3), the treated group is composed of students living more than 50 Km away from a university campus, and the control group of students living less than 20 Km from a university campus. Year2021 is a dummy variable that equals 1 for localities in 2021 and 0 otherwise. Year2021*$Campus_L$ takes the value of one for localities in 2021 from the treated group defined as explained before. The regression includes all localities with more than 2,000 inhabitants.