

DISTANCE LEARNING IN HIGHER EDUCATION: EVIDENCE FROM A RANDOMIZED EXPERIMENT

M. Paula Cacault

EPFL and Enterprise for Society (E4S)

Jérémy Laurent-Lucchetti

University of Geneva

Christian Hildebrand

University of St. Gallen

Michele Pellizzari

University of Geneva

Abstract

Using a randomized experiment in a public Swiss university, we study the impact of online live streaming of lectures on student achievement and attendance. We find that (i) attending lectures via live streaming lowers achievement for low-ability students and increases achievement for high-ability ones; (ii) students use the live streaming technology only occasionally, apparently when random events make attending in class too costly, and (iii) offering live streaming reduces in-class attendance only mildly. These findings have important implications for the effective design of education policies. (JEL: I20, I21, I23)

1. Introduction

In the past decades, almost all sectors of the economy have experienced enormous changes in production technologies. Education has been slow at joining this trend but the so-called *Edutech* industry is now booming with a large and growing offer

The editor in charge of this paper was Giovanni Peri.

Acknowledgments: This project has benefited from the help and support of many persons in the administration of the University of Geneva. We would like to thank Stéphan Berthet, Marcelo Olarreaga, Emmanuel Rousseaux, José Stultschnik, Maria-Pia Victoria-Feser, and Nicolas Zingg. We would also like to thank all the teachers who participated in the experiment: Jean-Paul Chaze, Giovanni Ferro-Luzzi, Diego Kuonen, Frederic Robert-Nicoud, Federica Sbergami, Thomas Straub, and Lammert Vrieling. Our gratitude also goes to all the students in the experimental classes. We received valuable comments from seminar participants at the University of Geneva, Universitat de les Illes Balears, EU-JRC (Ispra), Warwick CAGE Workshop in Applied Economics, the 2019 SKILS Workshop, ANVUR, 3rd Applied Microeconomics Workshop at the University of Bolzano, Queen Mary University of London, and Paris School of Economics. All errors remain our sole responsibility. The experiment presented in this paper is registered in the AEA RCT Registry with ID AEARCTR-0004107. Pellizzari is a Research Fellow at the CEPR and IZA.

E-mail: Paula.Cacault@epfl.ch (Cacault); Christian.Hildebrand@unisg.ch (Hildebrand); Jeremy.Lucchetti@unige.ch (Laurent-Lucchetti); Michele.Pellizzari@unige.ch (Pellizzari)

Journal of the European Economic Association 2021 19(4):2322–2372 DOI: 10.1093/jeea/jvaa060
© The Author(s) 2021. Published by Oxford University Press on behalf of European Economic Association.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

of e-learning platforms, distance learning tools, and massive open online courses (MOOCs). Given the enormous increase in the demand for higher education around the world, distance learning technologies are attracting special attention given their cost efficiency (Deming et al. 2015) and the crucial role they played during the COVID-19 pandemics. The academic literature exploring the effects of these new technologies on students' outcomes is also expanding fast and so far the evidence suggests that results depend heavily on the institutional context (U.S. Department of Education 2010; Escueta et al. 2017).

In this paper, we provide novel experimental evidence on the impact of a distance learning technology on student performance and classroom attendance in a standard European public university. We conducted a large-scale randomized experiment in which first-year bachelor students were offered access to a live streaming platform for many of their compulsory courses. Access to the platform was randomized both across students and over weeks of the term, so that the same student could attend the classes online in some weeks but not others. Furthermore, we constructed a mapping of all questions in the final exams to the weeks of the term in which the material required to answer them was covered. Therefore, we can exploit variation both across and within students and weeks. The students in our experiment were not forced to attend classes online and they could always decide to attend in person. This adds a layer of complication to our analysis but also allows investigating students' choices of attendance modes.

We believe that our study provides an important contribution to understanding how distance learning technologies affect attendance and achievement in higher education. To the best of our knowledge, we are the first to implement a research design using experimental variation across students and exam questions. This research design allows investigating heterogeneity across the distribution of students' ability, something that has rarely been possible in previous studies. In addition, we conduct our study in a context that is representative of many higher education institutions around the world and we also cover a broad set of lectures. This contrasts with many existing studies that focused either on private for-profit colleges or on a very limited set of classes and subjects. As a result, we believe that our results generalize easily to a large number of other interesting institutional contexts.

The students in our experiment can watch lectures online only in some randomly selected weeks of the term, whereas the rest of the time they only have access to classroom lectures. All other features of the course logistics are typical of a traditional teaching environment. Thus, according to the taxonomy proposed by U.S. Department of Education (2010), our setting employs a *blended* or *hybrid* learning approach to teaching, where the use of online technologies does not replace but rather complement other teaching practices. Our experiment provides an ideal opportunity to identify the causal effect of what is arguably the most salient feature distinguishing online from offline education, namely, the fact that the material is presented to the students via the screen of an internet connected device located at any distance from the teacher. In our experiment, this is the only difference between the treatment and the control conditions.

For the same reason, our results cannot be interpreted as identifying the causal effect of fully online against fully offline education, because students always have the option to attend lectures in the classroom or meet the instructors in person. The

few studies comparing fully online with fully offline education programs have found results that are extremely dependent on the specific features of such programs (Escueta et al. 2017). Thus, it seems crucial to investigate the role of each such features in details and our study focuses on the single most meaningful of them, namely, the mode of knowledge transmission being online or face-to-face.

Our analysis shows that students use live streaming only occasionally, apparently when attending in class is too costly (e.g., due to shocks to health or commuting conditions), and that attending lectures online has a negative impact on achievement for low-ability students and a positive one for high-ability ones. Having access to the streaming platform (i.e., the intention-to-treat [ITT] effect) reduces the probability of answering exam questions correctly by 2 percentage points for students in the lowest quintile of the ability distribution (over an average probability of 55%).¹ The effect at the opposite end of the ability distribution (i.e., students in the top 20%) is positive and equal to an increase of 2.5 percentage points. Additionally, we find that offering access to the live streaming platform has only a modest effect on class size: on average, only 8 out of 100 students with access to live streaming do not show up in class. Apparently, when given the choice, most student prefer to attend lectures in class. This is an important finding that can guide the design of education policies across institutional contexts. For example, simple live streaming technologies are unlikely to offer a solution to institutions facing overcrowding of their classrooms.

We discuss plausible mechanisms that might rationalize these findings. In particular, the data seem consistent with the idea that students prefer attending classes in person but use the streaming service when random events make the cost of in-class attendance too high (e.g., sickness and bad weather). To rationalise the heterogeneous effects on performance, we explore the possibility that the counterfactual to streaming varies across the distribution of students' ability. In large classes, like the ones in our experiment, interactions with the teacher are limited and the benefits of attending the class might be only mildly lower than that of in-class attendance, especially for the very good students who can presumably understand a lot of the material on their own. It is therefore possible that students at the upper end of the ability distribution stream instead of not attending when hit by an attendance shock. The streaming service for them is akin to an "insurance" against the risk of not being able to attend. By contrast, students of lower ability might find it harder to understand the material without the explanations given in class, implying that own study is relatively inefficient compared to attendance in class. Hence, without streaming, they tend to go to class to maintain their level of understanding even when the attendance cost is high. They use streaming, when available, instead of attending as it avoids the cost of attendance and only mildly reduces their grades (thus bringing net welfare gains). Under this view, the counterfactual to streaming might be no attendance for the good students and regular in-class attendance for the low ability students. Although we cannot pin down this specific mechanism with the data available to us, we inform the discussion

1. We measure ability using high school grades.

with novel evidence about attendance patterns across both types of students and types of classes that we obtain from an follow-up survey of our students.

We believe that our findings have important policy implications. The large and heterogeneous effects across students of different abilities is a finding that can readily inform the design of education policies related to the use of distance learning technologies. For example, university administrators who care about the academic performance of their students learn from our analysis that simply streaming classroom lectures might increase the learning gap between high-ability and low-ability students. It is difficult to infer from our results much about the effects of more sophisticated and better structured (and, presumably, more expensive) technologies for distance learning. However, our paper should at least warn about the potential negative effects on learning for the least able students and encourage institutions to carefully design, and possibly test, such new technologies. Beyond its effect on student achievement, universities might still want to offer live-streaming services as part of the amenities contributing to the consumption value of tertiary education that has been documented to be important for attracting students (Jacob et al. April 2018). In fact, our theoretical discussion highlights that introducing streaming unambiguously increases welfare for all students. Even low-ability students are made better off when they can live stream lectures, despite the negative effect on grades. Beyond its cost effectiveness, live streaming could thus also be of interest to universities as a tool for attracting new students.

We contribute to a growing literature on *Edutech* and on distance learning in particular (U.S. Department of Education 2010; Escueta et al. 2017). Despite the enthusiasm often surrounding these innovations, the available studies suggest caution. For example, McPherson and Bacow (2015) and Banerjee and Duflo (2014) document high drop-out rates for MOOCs and Bettinger et al. (2017) show worse labour market outcomes for graduates of online for-profit colleges and document a pattern of heterogeneous effects suggesting a lower negative effect for the most able students. Importantly, this pattern is perfectly consistent with our findings. Similarly, Deming et al. (2016) show that online for-profit colleges are viewed less favourably by employers than brick-and-mortar schools of all types. Brown and Liedholm (2002) show suggestive evidence of substantially lower achievement for college students attending classes in mixed or fully online modes. Swoboda and Feiler (2016) and Setren et al. (2020) focus on a teaching approach known as flipped classes, where students watch pre-registered videos of the lectures and classroom time is used to discuss problem sets with the instructors. Consistent with our study, these papers also find effects on learning outcomes that are increasing with the students ability.

While there exist a myriad of descriptive studies (see U.S. Department of Education (2010) for a review), few papers provide plausibly causal estimates of the use of distance learning tools in higher education. For example, Bettinger et al. (2017), Coates et al. (2004) and Xu and Jaggars (2013) use instrumental variables. All these studies exploit some version of commuting time from the students' residence to the college as an instrument for the take-up of distant learning but they differ in many other dimensions. Coates et al. (2004) is a small-scale study of one course in Introduction to Microeconomics involving 127 students, which finds small positive effects.

Bettinger et al. (2017) instead exploits a large sample of students from one large multi-campus for-profit institution and find negative effects on achievement. Xu and Jaggars (2013) looks at a large sample of students at community colleges and finds robust negative effects.²

A few other studies provide small-scale experimental evidence.³ Alpert, Couch, and Harmon (2016), Figlio, Rush, and Yin (2013), and Joyce et al. (2015) have samples of 300–700 students in introductory microeconomics classes and find mixed results. Bowen et al. (2014) have a sample of 600 students across 6 institutions and show no effects. Participation in all these experiments was incentivized, thus making it difficult to investigate the take-up decision.⁴

Our paper also speaks to the literature on for-profit colleges, which make extensive use of online learning technologies but mostly attract students of relatively lower ability compared to more traditional institutions (Deming, Goldin, and Katz 2012; Deming et al. 2013). Our findings suggest that this education model might in fact be more effective on higher ability students, with the important caveat that for some students the counterfactual to enrolling in a for-profit college might simply be less education.

As for any experimental study, our results may not easily generalize to other settings. Nonetheless, we are confident that our environment is quite similar to those of many public or private non-for-profit institutions of higher education, which are still prevalent in most countries around the world.

The paper is organized as follows. In Section 2, we describe the institutional setting and the experimental setup. Section 3 describes the data that we use for the empirical analysis, which we introduce and discuss in Section 4. Section 5 presents a discussion on mechanisms that is useful to interpret the empirical findings and section 6 concludes.

2. Institutional and Experimental Setup

2.1. General Context

As part of a general effort toward more intensive use of Edutech tools, in 2016 the University of Geneva was considering the implementation of an online platform to offer students the possibility to watch the lectures remotely from internet-connected devices. Given the cost of setting up such a system on a large scale and the concerns with its potential effects on the quality of learning, it was decided to first test it on an experimental basis.

2. The same authors find similar results in a previous companion paper using propensity score matching and restricting the sample to courses in mathematics and English (Xu and Jaggars 2011).

3. Oreopoulos and Petronijevic (2018) also presents results from a randomized experiment using technology in higher education but they focus more on learning support services, such as coaching and online exercises.

4. In Alpert, Couch, and Harmon (2016) the incentive was a five-point bonus on the course average grade.

We decided to target classes in the bachelor program in economics and management (EM), one of the most popular programs offered by the university and one where the benefits of distance learning could be high, given the overcrowding of some classrooms, especially in the first year. This bachelor is organized in two parts. The first part consists of the initial three semesters with a fixed sequence of compulsory classes, covering the foundations of economics, management, mathematics, and statistics. At the end of the third semester, students choose to major either in economics or management. Within each major there are both compulsory and elective courses for another three semesters. All the compulsory courses of the first part are taught in (at least) two parallel sections, one in English and one in French.

We chose to focus on a live streaming service as opposed to a simpler system of recording of the lectures. This decision was motivated by the observation that the latter, which had been available at the university since a few years already, did not appear to have any visible effect on attendance.

The system that we implemented consists of live broadcasting lectures taking place in a regular classroom environment. The live video streams can be accessed via a secured web platform, which is accessible only to the students who were assigned to treatment. The video lectures are not recorded and can only be viewed live, during the exact same time as the physical lecture taking place in the classroom.⁵

2.2. Experiment Setup

During two academic semesters, Spring 2017 (February–May) and Fall 2017 (September–December), we live-streamed all the lectures of eight compulsory courses of the bachelor program. We chose these courses as they were taking place in the largest auditorium of the school, the only one we could equip with the necessary technology.⁶ The official maximum capacity of the auditorium is 450 seats. Only regular lectures and not assistants' sessions were streamed.⁷ Table 1 lists the participating courses, the language of instruction, the bachelor programs for which they were compulsory and the number of enrolled students.

The experiment involved the courses of *Introduction to Macroeconomics*, *Probability and Statistics* and *Human Resource Management* in the Spring semester 2017. All three classes were taught in French. Probability and statistics was compulsory not only for students in the EM bachelor but also for those in International Relations (IR) and this is the reason why it is the most numerous of the three courses taught in this semester.

5. Students could in principle record the screen of their computers. The provider of the streaming service guarantees that it is only possible by physically filming the screen with an external camera, which would be very inconvenient to implement and resulting in very low quality.

6. These were the courses that were scheduled to take place in the auditorium already before the experiment. We did not manipulate in any way the classroom assignment process.

7. Importantly, during the semesters of the experiment, students also attended other compulsory courses which were never streamed.

TABLE 1. Courses participating in the experiment.

Term	Course	Language	Students	Bachelor ^a
Spring	Introductory macroeconomics	French	386	EM
	Probability	French	490	EM&IR
	Human resource management	French	242	EM
Fall	Introduction to microeconomics	French	460	EM
	Introduction to microeconomics	French	357	IR
	Introduction to microeconomics	English	241	EM&IR
	Mathematics	French	481	EM&IR
	Introduction to management	French	261	EM

Notes: a. EM: Economics and management. IR: International relations.

The following term (Fall semester 2017), we experimented with three additional courses: *Introduction to Microeconomics*, *Mathematics*, and *Introduction to Management*. Introduction to microeconomics was taught in three parallel sections, two in French (one of which entirely dedicated to IR students) and one in English. Each section was taught by a different professor but all three were fully harmonized, that is, covered exactly the same content, used the same problem sets, and the same slides and lecture notes (only translated in the two languages). All students in the three sections took the exact same exam.⁸ Students were assigned a section at the beginning of the term and were not allowed to switch. IR students were all assigned to their dedicated French section, whereas EM students could choose the French or English section but were asked to stick to their choice for the entire term. Mathematics was also compulsory for both EM and IR students.

All these courses are organized according to the same model, with one lecture and one teaching assistant (TA) session per week, each of them taking place in a time slot of 2 hours, with 90 minutes of actual teaching. Usually the teaching sessions (both the lectures and the TAs) start at 15-past the hour, run for a first part of 45 minutes, allow a 15 minutes break, and run for another 45 minutes.

The students involved in the experiment in the two semesters are for the most part different students. In the Spring 2017, we mostly have first-year students (first enrolled in September 2016) attending their second semester of compulsory courses. In the Fall 2017, most students have just enrolled (in September 2017) and are attending their first semester of compulsory courses. A few students are present in both semesters.⁹

The streaming platform was accessible 10 minutes before the lecture started and until 10 minutes after the end. The video showed the lecturer and the projector screen. In addition it showed the screen capture of the classroom computer on an adjustable window. The stream video could be zoomed and frozen but it could not be recorded for later

8. Students can take the exam in the language of their choice.

9. These are students in the first cohort, that is, those who entered the university in September 2016, and who retook one or more courses of the first semester with the second cohort.

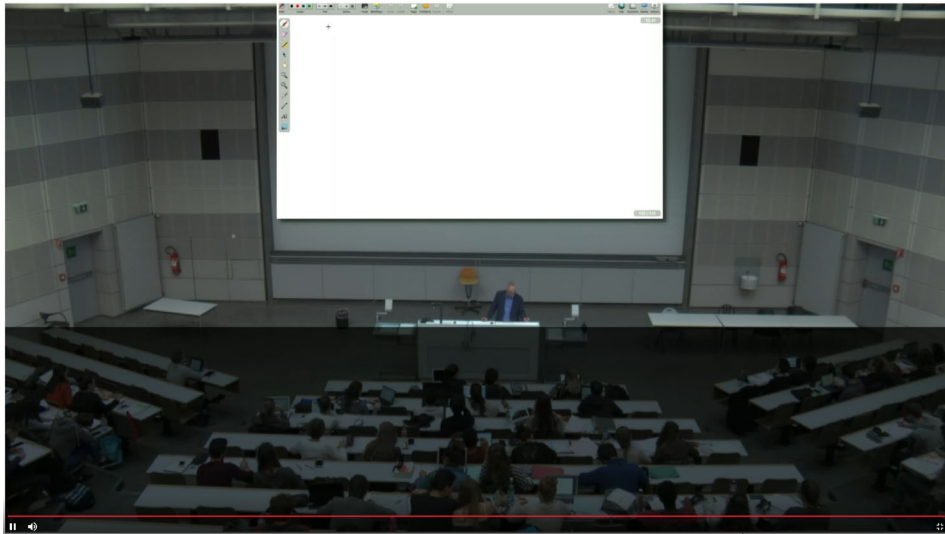


FIGURE 1. Live streaming of a classroom lecture. Screenshot of a lecture visualized via the live-streaming platform. The screen shows a view of the room from the back, the lecturer, and the projection of slides.

viewing.¹⁰ Figure 1 shows an example from the streaming service. The black band at the bottom was added to prevent seeing the computer screens of the students in the room.

Students accessed the streaming platform using their usual university credentials, that is, the same credentials needed to check the university email, to in courses, and exams. Hence, sharing access with other students was problematic, as it would also imply sharing access to all these other services.

At the University of Geneva, an academic term consists of 13 weeks (plus 1 week of mid-term break). In week 1, the professors of the participating courses presented the experiment in class to their students, who then had 2 weeks to enrol in the university's e-learning platform (they would have to do this regardless of the experiment as the e-learning platform is regularly used for sharing documents, announcements, submitting assignments, etc.). Based on the enrolment lists of each course from the e-learning platform, we first randomly assigned students to three groups. A first group of students (15% of all students) never had access to the streaming service and we label this group the *Never-access*. Another 15% of the students were given access to the service in all the weeks of the term and we label this group the *Always-access*. The remaining 70% of students were given access to streaming only some weeks at random and we label this group the *Sometimes-access*. Every week, a varying share of students in this group was given access. In the Spring semester 2017, we randomly

10. The provider of the streaming service is a leading company in this industry—livestream.com—and we subscribed to their most advanced platform plan (Enterprise), used also by large companies such as Spotify and USA Today.

assigned weekly access to 50% of the sometimes-access group. In the Fall semester 2017, we decided to vary this share between 20% and 80%.¹¹

At the end of week 2, students were notified by email about the exact sequence of weeks when they had access to the streaming platform.¹² Lectures started being streamed in week 3 and through the entire duration of the term. When a student had access to the streaming platform, she could watch the lectures of all the participating courses she was enrolled in for that week. Physical attendance in the classroom was always possible and students could freely decide to go to class in person even in the weeks when they had access to streaming.¹³

3. Data

We combine data from different sources. First, we collected information from the administrative records of all the students participating in the experiment and we observe their background characteristics such as age, gender, nationality (both of the individual and of their high school diploma), high school grade, and residential address.

Second, we collected detailed data from the streaming server allowing us to assess if, when, and for how long each student was logged in.¹⁴ Of course, we also know which students had access in which weeks.

Third, we merged individual-level data on student performance in the final exams. All the courses involved in the experiment had multiple choice exams, with the total

11. A total of 80% in week 3, 40% in week 4, 60% in week 5, 20% in week 6, 80% in week 7, 40% in week 8, week 9 was mid-term break, 60% in week 10, 20% in week 11, 80% in week 12, 40% in week 13, and 60% in week 14.

12. Figure A.1 in Online Appendix A shows an excerpt from the notification email that was sent to the students.

13. In order to smooth the implementation of the experiment and avoid resentment from those who had never or rarely access, we allowed students to request access for more specific weeks on an ad-hoc basis if they needed it for special reasons (sickness, travel, or other). The requests were assessed by the administrators of the experiment and the students were notified if they were accepted or not. In the end, we received very few such requests (two in the Spring and three in the Fall semester) and all were accepted. Including or excluding these observations from the sample has no meaningful effect on the results.

14. In fact, the information on the length of the connection is difficult to interpret, as log out can happen in a variety of different ways: actual log out, simply closing the lid of the laptop, closing the browser (or a tab of the browser), logging out of the operating system, switching off the computer, and so forth. Each of these events are recorded or not recorded depending on the machine (laptop, desktop, tablet, or smartphone), the operating system, and the browser. In addition, when recorded these events are coded in different ways and they are not always distinguishable from other events, such as activating the screen saver or switching to a different application. In Figure A.3 in Online Appendix A, we show the distribution of connection times (as percentage of the duration of the session) based on our best effort to detect log out events (and only for students who did connect to the platform). Despite the low quality of this information, the results suggest that a majority of students stay logged in for more than 80% of the session.

number of questions ranging between 16 and 60.¹⁵ For each student, each exam, and each question, we know whether a correct answer was provided. In addition, we asked all instructors to provide a detailed mapping of exam questions to weeks. For each exam question, the mapping indicates the week or weeks of the term when the material required to answer correctly was covered in the lectures. The mapping is not necessarily one-to-one, as some questions require material covered in multiple weeks and, reversely, the material covered in certain weeks is useful for multiple questions. There is, however, large variation allowing us to exploit differences in the patterns of correct answers across weeks when individual students had access to the streaming platform.¹⁶

Next, we have proxy measures of classroom attendance, unfortunately only for the Fall term 2017. We asked a research assistant to take multiple (two or more) pictures from the back of the auditorium, 15–20 minutes after the start of each lecture. Class attendance based on the pictures was evaluated by independent ratings of *Amazon Mechanical Turk* (MTurk) workers (see Figure A.2 in Online Appendix A for a sample of the pictures that were evaluated in this task). A total of 252 MTurk raters were randomly assigned to approximately 20 classroom sessions (in random order to avoid carry-over or sequence effects), each consisting of at least 2 pictures. Hence, each classroom session was evaluated by more than 100 MTurk workers. Raters were informed that the room could seat a maximum of 450 persons and they received a bonus payment based on the (absolute) deviation of each assessment from the (ex-post) arithmetic mean taken across all raters.

We also wanted to explore the idea that students use the streaming service only when hit by shocks that increase the cost of attending classes in person. To this end, we collected additional data on events that might plausibly affect such cost, like weather conditions and influenza epidemics. For each course and each week in our experiment, we have identified the exact calendar day when lectures took place and we have linked each of them to detailed weather variables from the *Swiss Federal Office of Meteorology and Climatology*. For each day, we have information on minimum, maximum, and average temperature, minimum and maximum barometric pressure, millimeters of rainfall, and minutes of sunshine in the Geneva region. We use these variables to code each day into one of three categories. “Bad weather” days are days with rain in the top quarter of the seasonal distribution, minimum pressure, and maximum temperature in the lowest quarters of the respective distributions.¹⁷ “Good weather” days are days with minutes of sunshine and maximum pressure in the top quarter of the seasonal distributions and with minimum temperature above the 25th percentile. Any day with neither “good” nor “bad” weather is coded as regular.¹⁸

15. Introduction to management has 60 questions, mathematics 16, introduction to microeconomics 30 (all sections), introduction to macro 30, probability and statistics 16, and human resource management has 30.

16. On average, the material covered in a particular week is useful to answer 2.5 questions.

17. We compute the seasonal distributions using measurements for all days from September 2016 to December 2017.

18. We have experimented with several variations of these definitions and results change only marginally.

In addition, the *Swiss Federal Office of Public Health* publishes weekly statistics about influenza incidence rates, where each week is coded from zero (no propagation) to three (extended propagation). Information is available at the detailed municipality level and it is based on doctor consultations.¹⁹ We have merged this information for Geneva with the weeks of the lectures in our experiment and we have coded days of outbreak any day with propagation above zero.

Finally, in order to investigate the potential mechanisms underlying our findings (see Section 5), we decided to collect additional information on attendance patterns for different types of students and classes. We designed an ad-hoc online survey with questions about attendance of compulsory and elective classes and also with a simple choice experiment where students were presented with pictures of crowded or non-crowded classrooms and were asked to indicate how likely they were to attend the class in person or online. The survey was anonymous (hence, we cannot link responses with administrative data) but contained questions about high school grades (together with various other background characteristics of the students) allowing us to measure ability in the same way as in our experimental data. We administered the survey in April–May 2020 to all students enrolled in one of the programs (bachelor and master) offered by the Faculty of Economics and Management at the University of Geneva. A total of 1,209 students were invited to the survey and 608 responded. With perhaps only a few exceptions, the respondents to the survey did not participate in our main experiment, which was run in 2016–2017. Nevertheless, there were no major changes in academic programs nor in admission procedures between 2017 and 2020, hence we do not expect the students in our experiment to be significantly different from the survey respondents.²⁰ We provide additional information about the survey and some descriptive statistics in Online Appendix D.

3.1. Descriptive Statistics

Our main dataset consists of 1,459 students, 459 of whom attended courses in the term of the Spring 2017 and 1,000 in the term of the Fall 2017. There are 162 students who are present in both terms. On average these students took 1.8 of the 8 courses participating in the experiment.

Table 2 shows descriptive statistics of the sample of students by assignment group. On average, students had access to the streaming service for about 5.6 weeks, with the expected variation across groups. The take-up of the service was rather low, averaging around 10% and going up to 12% for the always-access group.

19. Source: Office Fédéral de la santé publique, *Sentinella OFSP*.

20. The survey was administered during the COVID-19 lockdown, when all courses were being offered online and students spent a lot of their time in front of their computers. This presumably helped obtaining high response rates but might also affect somehow the answers to some questions. We were very careful at phrasing the questions as referring to a regular semester or class and to encourage students to abstract from the emergency situation when answering.

TABLE 2. Descriptive statistics by experimental condition.

	All		Never-access		Sometimes-access		Always-access	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Background								
Swiss = 1	0.601	0.490	0.612	0.488	0.602	0.490	0.584	0.494
Female = 1	0.507	0.500	0.570	0.496	0.492	0.500	0.513	0.501
Age in years	20.119	2.217	19.814	2.052	20.158	2.186	20.248	2.482
High school grade ^b	0.333	0.184	0.338	0.191	0.330	0.182	0.339	0.186
Swiss high school	0.560	0.497	0.540	0.499	0.566	0.496	0.556	0.498
French high school	0.300	0.459	0.298	0.458	0.299	0.458	0.308	0.463
Mother with college	0.352	0.478	0.367	0.483	0.355	0.479	0.328	0.470
Father with college	0.428	0.495	0.422	0.495	0.424	0.494	0.450	0.498
Treatment								
Weeks with access	5.626	3.350	0.000	0.000	5.693	1.631	11.000	0.000
Take-up rate ^a	0.104	0.202	.	.	0.100	0.199	0.120	0.211
Outcomes								
Course-drop ^c	0.295	0.456	0.286	0.453	0.295	0.456	0.308	0.463
Exam grade ^d	−0.003	0.939	0.047	0.870	−0.026	0.958	0.049	0.923
Share of correct answers ^e	0.556	0.204	0.568	0.197	0.553	0.204	0.556	0.211
Observations	1621		252		1119		250	

Notes: a. Share of students streaming at least once.

b. Average normalised high school leaving grade.

c. Share of students not taking at least one exam.

d. Average grade normalised within each course.

e. Share of correctly answered multiple-choice questions across all exams.

About 60% of the students are Swiss nationals and the rest are foreigners.²¹ About 50% are female and average age is 20 years old, with some limited variation generated by early and late enrollers and also by heterogeneity of schooling tracks across the nationalities of the students. As high school grades are recorded using different metrics in different countries, we normalized them using the formula:

$$\frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min}},$$

where x_{ij} is the grade obtained by student i in country j (in its original format), x_j^{\min} is the minimum passing grade in country j , and x_j^{\max} is the maximum obtainable grade in country j .²² Note that the normalized high school grade depends both on the range of possible grades in country j (the denominator of the formula) and the position of the student's grade within this range. A normalized high school grade of zero means that the student graduated with the minimum passing grade. A normalized high school

21. The second most frequent nationality is French. The University of Geneva has a very international student body, due to both the proximity with the French border and to the presence of a large community of expatriates.

22. Some countries changed the scales in recent years; we considered the system that was in place the year that the student obtained her diploma. In Germany, the schooling system uses a decreasing scale, so we reverted the scale to compute the normalized grades.

grade of 1 means that the student graduated with the maximum grade. On average, our students graduated with 33% of the difference between the maximum and the minimum passing grade.

Around 35% (43%) have a mother (father) with a college degree. Consistent with the random allocation, differences in all these descriptive statistics across the three assignment groups are minor and non-significant.²³

The lower panel of Table 2 reports descriptive statistics for some outcome variables. We define course-drop as the proportion of students who did not participate in at least one of the final exams of the courses they enrolled in. On average, 30% of the students drop a course. This is a high percentage and is the result of both students choosing to take some exams at a later stage as well as to some of them dropping out of university altogether.²⁴ We conducted a series of robustness checks and did not find any effect of the experimental conditions on course-drops (see Table A.2 in the Online Appendix).

Exam grades originally range on a scale of 1–6 points and we have standardized them to have mean zero and variance one within each course.²⁵ There are some differences across groups, with higher grades for the never-access and always-access students but none of these differences is statistically significant.

As we discussed above, we also have information about each exam question and we can compute the share of correctly answered questions for each student across all the exams they take. This is reported in the last row of Table 2. On average, students answer correctly around 55% of the questions and the differences across assignment groups are minor and not significant.²⁶

Comparing Table 2 and Online Appendix Table C.2 allows understanding the differences between the sample of students who participated in the main experiment and those who responded to our online survey. Especially when restricting the survey sample to first-year students, like those in our experiment, the two samples appear to be quite comparable, especially in terms of age and language.²⁷ The survey participants are more likely to be female (60% against 50%) and they display higher high school grades (0.44 against 0.33). These differences are presumably due to heterogeneity in response rates.

23. Table A.1 in the Online Appendix shows a full set of mean comparison tests for all variables and all pairs of assignment groups (30 tests in total). Only three such tests (marginally) reject the equality assumption.

24. In fact, dropping out is not uncommon in Geneva, as in many Swiss universities, where students with a valid Swiss high school diploma cannot be screened. As a consequence, many students enrol and drop out during their first year.

25. There are different numbers of students in each course, hence the grand average is not exactly equal to zero nor the standard deviation exactly equal to one.

26. Many questions have more than two possible answers.

27. Of the survey respondents, 86% completed the survey in French, which broadly corresponds to the sum of experimental students coming from either a Swiss or a French high school.

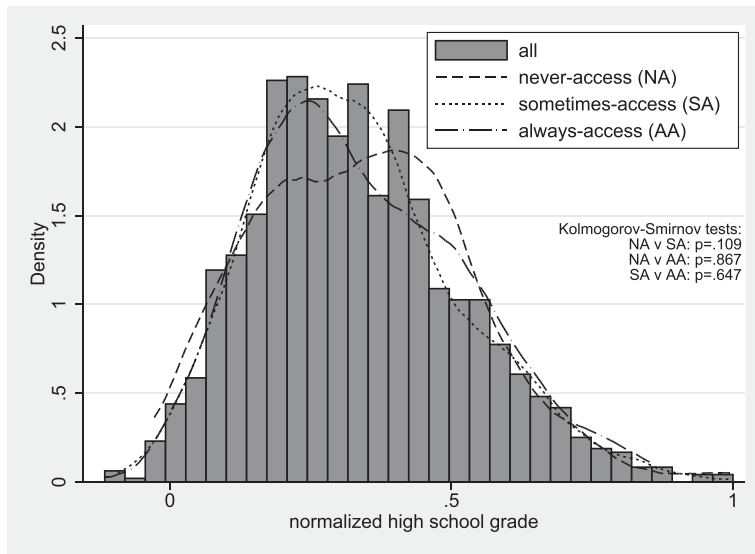


FIGURE 2. Ability distributions by assignment groups. The figure shows the estimated densities of normalized high school leaving grades, by assignment group. The note on the side presents p -values of tests for the hypotheses that distributions of grades of the two groups are equal.

In Figures 2 and 3, we provide further evidence that the randomization procedure was successful. Figure 2 plots the distribution of our preferred measure of student ability, namely their standardized high school leaving grade. We show the distribution for the entire population of students and also separately by assignment group and we report the results of formal Kolmogorov–Smirnov tests for the equality of the distributions across each combination of groups. Results confirm that the distributions are all similar across the entire range of the ability spectrum and across groups.

Randomly allocating students to one of the three assignment groups is only the first step of our randomization procedure. We further randomly assigned the students in the sometimes-access group to access the streaming service week by week. To check the quality of the overall procedure, we estimate a linear probability model using all the student-week observations (30,732) and with a dummy indicator as dependent variable which equals one if the student in the specific week was given access to the streaming platform. As explanatory variables we include a set of student characteristics: nationality, gender, cohort of enrollment, and nationality of high school diploma. We also add week-term fixed effects and we estimate the model with student random effects. Figure 3 shows the estimated coefficients of the student characteristics and their standard errors. Consistently with the random allocation, all these coefficients are small in magnitude and statistically indistinguishable from zero.

In Table 3, we report for each week of each term the share of students who randomly received access to the streaming platform, the share of those who accessed the platform and the corresponding take-up rates. On average, over the 11 weeks

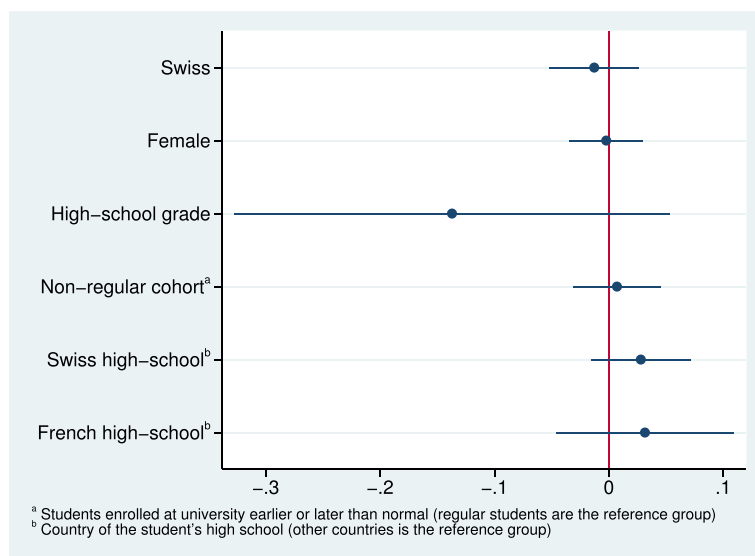


FIGURE 3. Randomization checks. The figure shows estimated coefficients and standard errors of a linear probability model, estimated on all student-week observations (30,732). The dependent variable is an indicator which equals one if the student in the specific week was given access to the streaming platform. The explanatory variables are nationality, gender, cohort of enrollment, nationality of high school diploma, week-term fixed effects and student random effects.

TABLE 3. Weekly assignment and take-up.

Week	Spring 2017			Fall 2017		
	Assigned ^a	Streamed ^b	Take-up ^c	Assigned ^a	Streamed ^b	Take-up ^c
3	0.497	0.059	0.119	0.679	0.052	0.077
4	0.525	0.057	0.108	0.425	0.049	0.115
5	0.484	0.040	0.083	0.556	0.066	0.119
6	0.549	0.058	0.106	0.284	0.026	0.092
7	0.503	0.042	0.083	0.700	0.082	0.118
8	0.471	0.031	0.067	Mid-term break		
9	Mid-term break			0.416	0.040	0.096
10	0.536	0.058	0.107	0.571	0.064	0.111
11	0.501	0.049	0.097	0.306	0.035	0.115
12	0.482	0.049	0.101	0.692	0.077	0.112
13	0.499	0.039	0.079	0.424	0.055	0.130
14	0.525	0.037	0.070	0.585	0.074	0.126
Total	0.507	0.047	0.093	0.513	0.056	0.110

Notes: a. Share of students with access to the live streaming platform.

b. Share of students who accessed the platform at least once (for any duration).

c. Share of students with access who streamed at least once.

TABLE 4. Proxy measures of classroom attendance.

	Enrolment ^a	MTurk evaluations ^b		
		Mean	Median	Std. dev.
Micro (FR1)	369	213.91 (61.96)	171.89 (21.87)	152.19
Micro (FR2)	513	210.36 (12.17)	182.67 (13.64)	124.27
Micro (EN)	249	116.69 (9.98)	89.29 (9.86)	90.64
Mathematics	486	193.89 (15.69)	167.70 (13.64)	124.40
Management	287	204.51 (15.11)	168.71 (12.31)	147.73

Notes: a. Number of students enrolled in the course.

b. Statistics computed over the 252 raters of the pictures of each classroom session. The standard deviations in parentheses are the average standard deviations of all the classroom sessions of the same course.

of each semester during which lectures were streamed, about 50% of the students had access to the platform.²⁸ Consistent with the experimental design, this share is relatively constant across the weeks of the Spring term, whereas it varies between 30% and 70% in the Fall term.²⁹

Using information from the server of the streaming platform, we can identify the students who actually accessed the service. On average, only about 5% of the students (i.e., including those who had no access) used the service at least once in each week. There is some variation in usage, which is slightly higher in the Fall than in the Spring term but with no clear pattern over the weeks.

Combining information on assignment and usage we also construct measures of take-up, that is, the share of students with access who logged into the platform. This share is on average around 10%–11%, ranging across weeks from a minimum of 6.7% to a maximum of 13%. It is also a little higher in the Fall than in the Spring term.

Finally, in Table 4, we report descriptive statistics for our proxies of class attendance. For each course, we show the total number of students enrolled (first column) and some statistics of the evaluations of the classroom pictures by the MTurk raters. More specifically, for each classroom session we have a distribution of the many evaluations by the MTurk raters. Hence, for each classroom session, we can compute, for example, the mean, the mode, and the standard deviation of such distribution. In Table 4, we report the average mean, mode, and standard deviations across all the classroom sessions of the same course.

Not all enrolled students always show up in class, hence the evaluations are on average lower than actual enrollment. The distributions of the evaluations also appear

28. Recall that these averages are taken over all three assignment groups.

29. In the Fall term, we varied the share of students with access week by week. See also footnote 11.

to be rather skewed to the right, with medians always lower than averages. In addition, there is also quite a bit of dispersion, with standard deviations ranging between about half of the average to very close to it.

4. Main Empirical Results

In this section, we investigate the effect of the experiment on three sets of outcomes: (i) take-up of the streaming service, (ii) classroom attendance, and (iii) exam performance.

4.1. *Take-up of the Streaming Service*

As we documented already in Section 3, on average, across the weeks of the experiment, only around 10% of the students who were given access to the streaming platform actually used the service.

This result might be generated by two very different types of behaviour. First, students might be heterogeneous in their interest towards this new service and only few of them like it enough to use it. In this case, we should see that the 10% of users are consistently the same persons across the weeks of the term. Alternatively, students might all have a preference to attend the courses in class but may prefer to use the streaming platform when random events make the cost of class attendance too high. We show here that the second mechanism is the most consistent with the data (and it is the one we develop in our discussion on mechanisms in Section 5).

Assuming that the probabilities of the attendance shocks are uncorrelated across time, the distribution of cumulative take-up (i.e., the share of times one uses the service if given access) should follow a Pareto-like distribution, with many students never taking up (because they are not hit by shocks) and a rapidly declining probability of taking-up more and more often. Only students who happen to be hit by a shock every time they have access to streaming should take up all the times.

Our data allow us to compute the distribution of cumulative take-up and we report it in Figure 4, both for the entire population of students who were given access at least once (left panel) and also restricting to only those who used the service at least once (right panel). We define cumulative take-up as the ratio between the number of weeks the student used the streaming service over the number of weeks she had access to the platform.

The figure clearly shows that the distribution of cumulative take-up follows a Pareto-like distribution, with the bulk of students never streaming and a quickly decreasing tail.

In Figure 5, we replicate the analysis by ability groups. We classify as “low-ability” students with high school grades in the lowest 20% of the distribution, “high-ability” students in the top 20%, and all the others as “mid-ability” students. The figure shows that there are no striking differences in the distribution of cumulative take-up across groups. In all cases, it clearly follows a similar distribution.

We also investigate whether weekly take-up varies with students ability. In Figure 6, we report the average take-up rate across weeks by percentiles of the

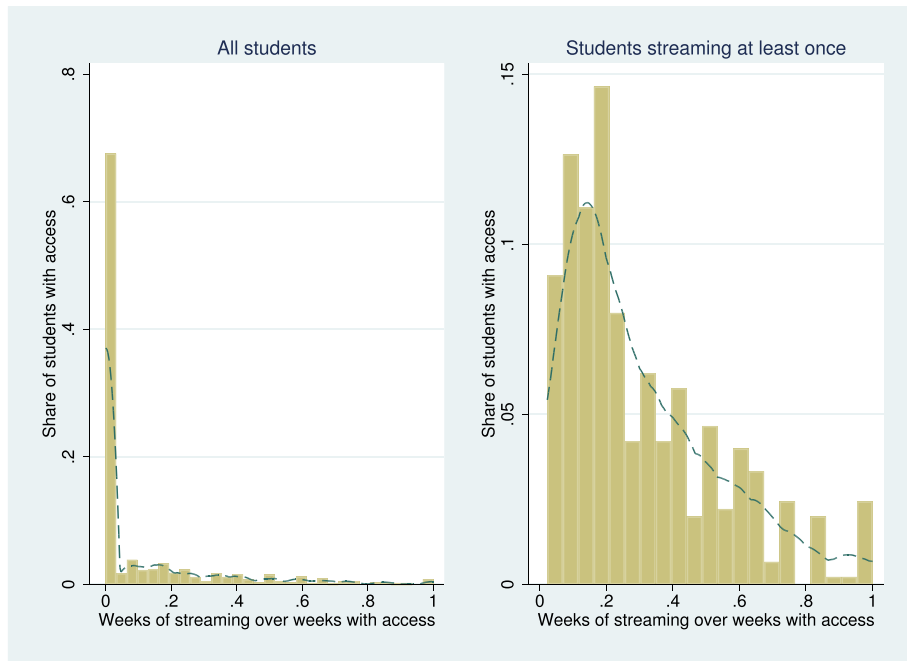


FIGURE 4. Distribution of cumulative take-up. The figure plots the distribution of cumulative take-up measured by the share of weeks in which the student connected to the platform over the total number of weeks with access. The right panel restricts the sample to students who used the platform at least once during the term.

distribution of high school grades. We compute take-up as the ratio between the number of students who actually logged into the streaming platform over the number of those who had access to it.

Despite some sizeable variation between 0% and 20%, there seems to be a pattern towards slightly lower take-up among the very low- and the very high-ability students. We highlight this pattern by reporting in the figure also the predicted take-up (and the corresponding 95% confidence interval) obtained from a simple probit model of weekly take up conditioning on a quadratic function of normalized high school grades and our standard set of control variables, which includes course and week fixed effects, a gender dummy, a dummy for Swiss nationals, dummies for enrolment cohorts (early, late and regular enrollers), and nationality of the high school diploma (Swiss, French, and other nationalities). The results of this estimation are reported in Table A.3 in the Online Appendix A.

4.2. Classroom Attendance

Given the low take-up rates, we expect the experiment to have only a limited impact on attendance.

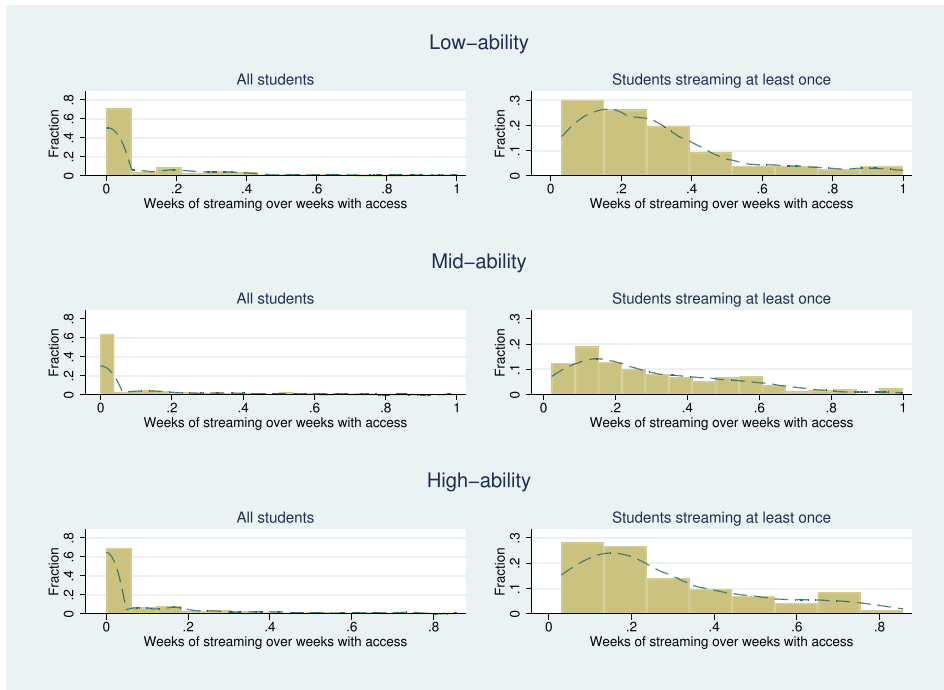


FIGURE 5. Distribution of cumulative take-up by ability. The figure replicates the analysis of Figure 4 by three broad ability groups. Low/mid/high ability groups are based on percentiles of the distribution of high school leaving grades. Low = 1st–20th, mid = 20th–80th, and high = 80th–100th

Students could always attend the lectures in person and presence in class is not recorded at the student level. Hence, the only information we can use for this analysis comes from the MTurk evaluations of the classroom pictures (see Section 3 for details). Unfortunately, this information is far from ideal. We only have 42 observations, 1 for each lecture participating in the experiment in the Fall 2017, and the MTurk evaluations are noisy, as documented in Table 4.³⁰

Eventually, we estimate regression models of the following type:

$$attendance_{cw} = a_1^1[\# assigned]_{cw} + a_c^1 + e_{cw}^1, \quad (1)$$

$$attendance_{cw} = a_1^2[\# streamed]_{cw} + a_c^2 + e_{cw}^2, \quad (2)$$

where the dependent variable is the median of the MTurk evaluations of the number of people attending the lecture of course c in week w ; $[\# assigned]_{cw}$ and $[\# streamed]_{cw}$

30. There were 55 lectures being streamed during the fall term, but unfortunately we were unable to take photographs of the auditoriums in 13 of them.

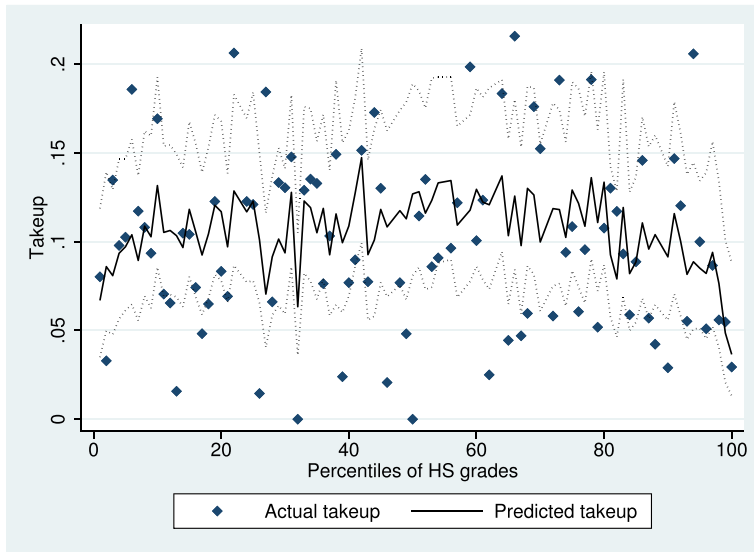


FIGURE 6. Distribution of weekly take-up by ability. The dots are the average take-up rate (the ratio between the number of those who actually logged into the streaming platform over the number of students who had access to it) by percentiles of the distribution of high school grades. The lines represent predicted take-up (and the corresponding 95% confidence interval) from a probit model of weekly take up conditional on a quadratic function of normalized high school grades, course and week fixed effects, gender, nationality, enrolment cohorts, and nationality of the high school diploma. The detailed probit estimates are reported in Table A.3 in the Online Appendix.

are, respectively, the number of students who were given access to the streaming platform and the number of students who actually logged in for course c and week w . Both models include course specific fixed effects (a_c^1 and a_c^2) and $\{e_{cw}^1, e_{cw}^2\}$ are residuals.

The parameters of interest are a_1^1 , which measures the ITT effect, and a_1^2 , which measures the average treatment effect on the treated (ATT), with treatment defined as using the streaming service. Given the potential endogeneity of students' take-up decisions, we estimate equation (2) using $[\# \text{ assigned}]_{cw}$ as an instrument for $[\# \text{ streamed}]_{cw}$ and by doing so, we clearly identify a local effect on the unobservable population of compliers. The first-stage coefficient is equal to 0.122 with a standard error of 0.013, suggesting that weak instrument concerns should not be a problem. In order to improve efficiency and given the large variance of the MTurk evaluations, we estimate both 1 and 2 by weighting observations with the inverse of the standard deviation of the evaluations of each course-week session.

The main results are reported in Table 5 and, despite the small sample size and the noisy measurements, they are quite reasonable. The ITT (column 1) is equal to -0.079 , indicating that for each student with access to the streaming platform the estimated number of people in the classroom pictures is 0.08 lower. Extrapolating, for every 100 students who are offered lectures via live streaming about 8 of them do not show up in class.

TABLE 5. Treatment effects on classroom attendance.

	ITT	OLS	ATT
# treated	-0.079** (0.037)	—	—
# streaming	—	-0.431 (0.260)	-0.654** (0.290)
Obs.	42	42	42
<i>Descriptive stat. of the dep. variable:</i>			
Mean	153.6	153.6	153.6
Std. dev.	37.1	37.1	37.1

Notes: Standard errors in parentheses. The dependent variable is the median evaluation of the classroom pictures by the MTurk evaluators. All columns include course fixed effects. All observations are weighted by the inverse of the standard deviation of the individual evaluations. Column (3) shows an IV estimate of classroom attendance using the number of assigned students as instrument for the number of students streaming.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The ATT is reported in the third column and shows that for every student actually logging into the streaming platform there are about 0.6 fewer students in the classroom (for every ten students logging into the platform about six of them do not show up in class). Following our results on take up, this ATT lower than 1 can be rationalized by the fact that for some students the counterfactual to streaming is not attending the classes. This mechanism would also explain why the estimated effect of streaming on attendance is lower than average take up (10%). Interestingly, the IV estimates are larger than ordinary least squares (OLS) (although not significantly) suggesting that students might stream more when the class is more crowded.

The results in Table 5 are novel and important. They suggest that, when given the choice between attending in class or online, only a few students choose the latter.

4.3. Grades

In this section, we present results of the effects of the experiment on exam performance. Our preferred specification exploits all the experimental variation available in our design, namely variation across both students and weeks of the term.

For each student i , in course c and week w , we define as dependent variable y_{icw} as the share of correct answers to the exam(s) question(s) related to the material covered in that course-week. The question-week mapping that professors produced for us is the crucial source of information for this analysis (see Section 3 for details). If only one question covered the material of the week, y_{icw} is binary (equal to 1 if the answer is correct and 0 otherwise). If more than one question in an exam referred to the material covered in a particular week, y_{icw} is the number of correct answers over the number of questions. Questions that require information from multiple weeks add to the grade of *each* of those weeks.

We use this dependent variable in linear models with the following specifications:

$$y_{icw} = \alpha_1^1 [\text{assigned}]_{iw} + \alpha_2^1 X_i + \alpha_c^1 + \theta_w^1 + \eta_i^1 + \varepsilon_{icw}^1, \quad (3)$$

$$y_{icw} = \alpha_1^2 [\text{streamed}]_{icw} + \alpha_2^2 X_i + \alpha_c^2 + \theta_w^2 + \eta_i^2 + \varepsilon_{icw}^2, \quad (4)$$

where $[\text{assigned}]_{iw}$ is a dummy indicator equal to one if student i was assigned to have access to the streaming platform in week w . Similarly, $[\text{streamed}]_{icw}$ is a dummy indicator for whether student i actually accessed the platform for course c on week w . X_i is a set of student controls including a gender dummy, a dummy for Swiss nationals, the normalized high school leaving grade, dummies for the three ability groups, dummies for three enrolment cohorts (early, late, and regular enrollers), and three nationalities of the high school diploma (Swiss, French, and other nationalities).³¹ $\{\alpha_c^1, \alpha_c^2\}$ and $\{\theta_c^1, \theta_c^2\}$ are course and week fixed effects, respectively. $\{\eta_i^1, \eta_i^2\}$ are individual effects that we treat either as fixed or random, depending on the specification. Hence, the observations referring to the same student are always allowed to be correlated with one another, both within and across courses. Obviously, when we estimate equations (3) and (4) with student fixed effects we cannot identify the parameters $\{\alpha_2^1, \alpha_2^2\}$. Finally, $\{\varepsilon_{icw}^1, \varepsilon_{icw}^2\}$ are residuals.

As with equation (2), we estimate equation (4) using $[\text{assigned}]_{icw}$ as an instrument for $[\text{streamed}]_{icw}$.³² The first-stage coefficient is 0.123 with a standard error of 0.003 (when we estimate the model on the sample of all students).³³

Results of the estimation of equations (3) and (4) are reported in Table 6. The upper panel shows the estimated coefficients for the entire population of students in our experiment, whereas the lower panel presents effects estimated separately for our usual three ability groups.³⁴

Results indicate that on average there is no detectable effect of the experimental assignment, nor of actual usage of the streaming platform. However, once we look at effects by ability groups, we uncover large heterogeneity, with a sizeable negative ITT effect on the low-ability students and a positive effect on the high-ability students.

31. The ability groups are defined as in Figure 6, on the basis of the distribution of the normalized high school leaving grade. Low-ability students are those with grades in the lowest 20% of the distribution, high-ability students those in the top 20%, and all the others as mid-ability students.

32. The treatment effect might in principle combine both the direct effect of watching the streamed lectures and also other effects, such as that of saving time from commuting and devoting it to studying.

33. Full results of the first stage regressions, including those estimated on the three ability sub-groups of students, are reported in Table A.4 in Online Appendix A.

34. These heterogeneous effects are obtained by interacting the variables $[\text{assigned}]_{icw}$ and $[\text{streamed}]_{icw}$ (both in the main regressions and in the first-stage of the IV model) with dummies for the three ability groups.

TABLE 6. Treatment effects on the share of correct answers.

	Random effects			Fixed effects		
	ITT ^a	ATT ^b	OLS ^c	ITT ^a	ATT ^b	OLS ^c
All students	0.001 (0.005)	0.006 (0.042)	−0.006 (0.011)	0.003 (0.005)	0.023 (0.043)	−0.005 (0.010)
<i>By ability group^d</i>						
Low	−0.019** (0.009)	−0.179* (0.098)	−0.020 (0.027)	−0.019** (0.009)	−0.178* (0.102)	−0.024 (0.027)
Mid	−0.000 (0.006)	−0.002 (0.047)	−0.003 (0.013)	0.000 (0.006)	0.001 (0.050)	−0.005 (0.013)
High	0.025** (0.010)	0.245** (0.111)	−0.011 (0.026)	0.023** (0.011)	0.241** (0.121)	−0.011 (0.028)
Obs. ^e	23766	23766	23766	23766	23766	23766
Mean and Std. dev. (in parentheses) of the dependent variable:						
All students	0.545 (0.378)	0.545 (0.378)	0.545 (0.378)	0.545 (0.378)	0.545 (0.378)	0.545 (0.378)
<i>By ability group^d</i>						
Low	0.505 (0.374)	0.505 (0.374)	0.505 (0.374)	0.505 (0.374)	0.505 (0.374)	0.505 (0.374)
Mid	0.527 (0.376)	0.527 (0.376)	0.527 (0.376)	0.527 (0.376)	0.527 (0.376)	0.527 (0.376)
High	0.653 (0.368)	0.653 (0.368)	0.653 (0.368)	0.653 (0.368)	0.653 (0.368)	0.653 (0.368)

Notes: Robust standard errors in parentheses. All regressions include the following controls: course fixed effects, week-wave fixed effects, a gender dummy, a dummy for Swiss nationals, dummies for enrollment cohorts (early, late, and regular enrollers), nationality of the high school diploma (Swiss, French, and other nationalities), and normalized high school leaving grade (only in the panel *all students*).

a. Columns named ITT show OLS estimates of the share of correct answers on assignment status (=1 if student has access in the given week for the given course).

b. Columns named ATT show IV estimates of the share of correct answers using assignment status (=1 if student has access in the given week for the given course) as instrument for streaming take-up (=1 if student connected to streaming in the given week for the given course).

c. Columns named OLS show OLS estimates of the share of correct answers on streaming take-up (=1 if student connected to streaming in the given week for the given course).

d. Low/mid/high based on percentiles of the distribution of high school grades. Low = 1st–20th, mid = 20th–80th, and high = 80th–100th. The estimates reported correspond to the coefficients on the interactions of the ability group indicators with the assignment status (column ITT), the instrumented streaming take-up (column ATT), or the streaming take-up indicator (column OLS).

e. One observation per student-course-week.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The same pattern of negative effects on low-ability students and positive effect on the high-ability ones remains for average-treatment-effects. Consistent with the experimental design, results are very similar regardless of whether we estimate the models with student random or fixed effects.³⁵

35. This result is also robust to alternative definitions of ability groups. Table A.5 in Online Appendix A replicates the bottom part of Table 6 and adds two panels with estimates of treatment effects using alternative partitions of ability percentiles.

The magnitudes of these estimates are sizeable. For students in the bottom 20% of the ability distribution, having access to the streaming platform (regardless of whether one uses it or not) lowers the share of correct answers by approximately 2 percentage points over an average of about 55%. The positive effect at the top of the ability distribution is even larger and in the order of about 2.5 percentage points. The ATT estimates are very large: around -18 percentage points for the low-ability and about $+25$ percentage points for the high-ability students.

To put these effects into perspective, it must be noted that the streaming service is only used with low probability, as we document in Section 4.1. Hence, if the streaming service was scaled up and offered to all students for the entire term, we should expect an increase in the share of correct answers equal to the ATT multiplied by the probability of take-up, which is around 10%, with little variation across ability groups. Hence, a decline of approximately 2 percentage points for low-ability students and an increase of about 2.5 percentage points for the high-ability ones. These effects coincide with the ITT, consistently with the idea that take-up is determined by random events. Considering that the mode number of questions in an exam is 30, these effects translate into about 0.6 additional wrong answer for the low-ability students and 0.75 additional correct answers for the high-ability students. Overall the exams in our experiment consist of a total of 182 questions and our estimates imply that the availability of the streaming service produces 3.5 additional wrong answers for low-ability students and 4.5 additional correct answers for the high-ability students.

The OLS estimates are always negative but rather small and never statistically significant. Our interpretation of the difference between the OLS and the ITT or ATT results is that students are apparently more likely to use the streaming service when the lectures cover material that is more difficult. In these cases, they are generally more likely to give wrong answers to the exam questions, which explains the negative OLS coefficients.

Figure 7 further explores the heterogeneity of the treatment effects by student ability. The figure reports the estimated ITT (left panel) and ATT (right panel) from regression models similar to equation (3) and (4) where the variables $[assigned]_{icw}$ and $[streamed]_{icw}$ are interacted with a full set of dummies for the student's quintile in the distribution of the high school leaving grades.³⁶ The ability groups are now rather small and we detect effects that reach the conventional levels of statistical significance only rarely. However, the figure clearly shows an evident upward trend in these effects, suggesting that the heterogeneity we documented in our main Table 6 is not a mere statistical artefact.

It is worth noticing that Figure 7 is remarkably similar to Figure 1 in Bettinger et al. (2017), only shifted upward along the vertical axis. Despite the numerous differences in the institutional contexts and identification strategies, both studies find that the effect of distance learning technologies increases with the student's ability. It

36. The estimates (and confidence intervals) in Figure 7 are produced with the specification using student random effects. Using fixed effects does not change the results in any meaningful sense.

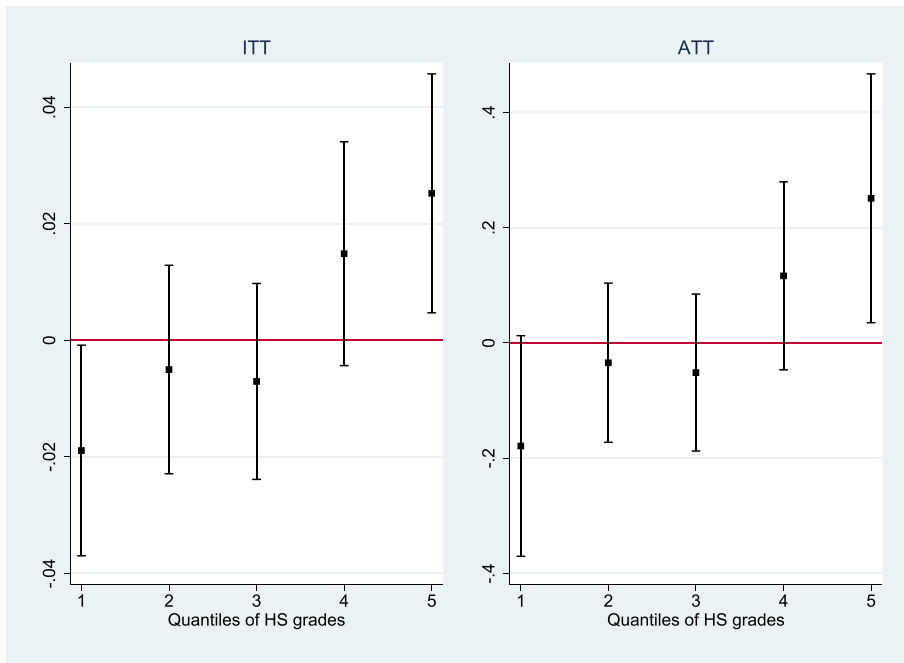


FIGURE 7. Treatment effects on the share of correct answers by ability. The figure reports estimates and 95% confidence intervals of the ITT (left panel) and the ATT (right panel) from regression models similar to equations (3) and (4), respectively, where the variables $[assigned]_{icw}$ and $[streamed]_{icw}$ are interacted with a full set of dummies for the student's quintile in the distribution of the high school leaving grades.

also seems natural to think that our analysis is based on a population of more skilled students compared to Bettinger et al. (2017), who focus on a for-profit college in the United States. Hence, they did not observe students in the upper part of the ability distribution for whom the effect of online learning turns positive. In Online Appendix B, we provide further evidence of that our results are consistent with Bettinger et al. (2017) by fully replicating their identification strategy with our data.

5. Discussion and Interpretation of Results

In this section, we discuss possible mechanisms that might rationalise the results we observe in the data.

We first consider the distribution of take-up. Figures 4 and 5 clearly show that cumulative take-up follows a Pareto-like distribution. This seems consistent with the idea that students use the service only when random events make the cost of class attendance particularly high.

To further explore this intuition, we have merged our data with information on weather conditions and influenza incidence rates in Geneva (see Section 3 for more

TABLE 7. Take-up and shocks.

Ability ^a	Daily weather			Influenza outbreak	
	Bad	Normal	Good	No	Yes
Low	0.142 (0.030)	0.091 (0.012)	0.117 (0.026)	0.090 (0.012)	0.130 (0.024)
Mid	0.213 (0.021)	0.109 (0.008)	0.170 (0.018)	0.113 (0.008)	0.156 (0.017)
High	0.161 (0.037)	0.084 (0.012)	0.106 (0.023)	0.089 (0.013)	0.087 (0.019)
All	0.189 (0.016)	0.101 (0.006)	0.147 (0.013)	0.104 (0.006)	0.138 (0.012)

Notes: Predicted probabilities of take-up based on estimates described in the text. Standard errors in parentheses.

a. Low/mid/high based on percentiles of the distribution of high school grades. Low = 1st–20th, mid = 20th–80th, and high = 80th–100th.

details). Although weather and health shocks do not exhaust all possible events affecting the cost of attending lectures in person, they certainly represent a meaningful set of such events. Weather, in particular, can affect both the conditions of the traffic in the city, and therefore the time required to commute, but also the opportunity cost of going to classes when the weather is good.

We have augmented the probit model for take-up with this merged data using student-week observations and adding the full set of three-way interactions between our usual ability groups, the dummies for daily weather conditions (good, bad, and regular days) and influenza epidemics (propagation or no-propagation).³⁷

For readability, we report in Table 7 predicted probabilities of take-up based on the estimates of this probit model, for different types of students and different weather and influenza conditions. We observe that students are almost twice as likely to log into the streaming platform on bad days than on regular days. This effect is present for students of all ability levels, with no clear differences between them. Good days and influenza also increase take-up but the effect is smaller than that of bad weather conditions.

Overall, the evidence on take-up seems consistent with the idea that students stream lectures only occasionally. Apparently, they normally prefer attending in class and use live streaming only as a replacement when the cost of going to class happens to be too large.³⁸ We believe that this is an important finding for understanding the potential impact of distance learning technologies and for the design of related policies.

37. We only modified the set of controls by replacing the week fixed-effects with season fixed-effects. The weather variables show too little variation within weeks to allow separate identification of the week fixed-effects.

38. It is noteworthy that the University of Geneva in 2018 extended the streaming service to other faculties (Law and Medicine especially). The data at our disposal show that the streaming usage is very similar to what we observed in our experiment: around 8%–10% of the registered students use the streaming service on average across all classes in other faculties as well.

Next, we turn to the heterogeneous effects on performance that we document in 6 and also in 7. The negative impact of streaming for the least able students can be easily rationalised with the rather natural idea that attending classes in person is the most effective learning technology. It is harder to interpret the positive impact that we detect towards the top of the ability distribution. One possibility could be that such heterogeneity emerges as a result of different counterfactuals.³⁹

Consider a situation in which the streaming technology is not available. Assume that in normal times most students attend lectures in person, but when the cost of going to class is too high, the good students tend to stay at home and study on their own. This happens because very good students can read the material in the book and understand most of it easily, even without the professors presenting and explaining it, whereas students of lower ability would have a harder time learning in autonomy and prefer attending.

Introducing the streaming technology in this context would allow all students to use it when the cost of going to class happens to be high. The good students replace own study with streaming, which improves learning and leads to the observed positive effect on exam performance. The low-ability students watch the streamed lectures instead of going to class, which is a more effective (but also more costly) mode of learning, and eventually perform worse. The students in the middle range of the ability distribution, substitute streaming to attendance for small shocks and streaming to no-attendance for larger shocks. Hence, average effect on grades tend to be close to zero, as we see in our data.⁴⁰

This mechanism has the potential to rationalise both low take-up and heterogeneous performance and it rests on the idea that the learning returns to attending classes are larger for the least able students. We believe that this is a reasonable assumption in the setting that we study, where classes are large and student–teacher interactions are limited, but it may not generalise to other settings. Most notably, in smaller classes with closer interactions, complementarities between student and teacher abilities would presumably emerge and make classes more beneficial to the more able students.

One implication of the mechanism we outlined here is that, in contexts where interactions with the teacher are limited, high-ability students are less likely to attend classes, especially when the cost of going to class is high. This result might appear somewhat counter-intuitive because of the common belief that good students are generally more likely to attend lectures.

In order to produce evidence on attendance patterns, we use data from the online survey that we described already in Section 3. In the survey we ask students the percentage of lectures they normally attend in person in a typical semester and we break

39. In Online Appendix C, we present a simple theoretical model that formalizes the mechanism we intuitively describe in this section.

40. It is possible that some of the shocks that trigger take-up of the streaming service might also have a direct effect of performance. For example, if a student is sick, this may reflect both in his/her propensity to attend lectures in person as well as to his/her ability to study and understand the material. For take-up induced by shocks of this type, we identify the joint effect of streaming and the shock itself (possibly interacted).

down the question for different types of courses. We then correlate these responses with ability, measured exactly as in our experiment, that is, high school grade rescaled by the minimum (passing) and maximum grades in the student's school system. Table 8 reports the results of this analysis, which we conducted by estimating simple OLS regressions with the reported share of attended lectures as dependent variable and ability, either linear (columns (1) and (2)) or in our usual three groups (columns (3) and (4)), as main explanatory factor. In addition, we experiment with and without a rich set of controls.⁴¹

Although sample sizes are somewhat small and the coefficients do not always reach conventional levels of statistical significance, the results clearly indicate that bachelor students with higher high school grades tend to report a lower percentage of classes attended in person. This is more evident in panel A, where we restrict the analysis to first-year students, like the ones in our main experiment, whose classes are significantly larger.⁴² Students in the top ability group report attendance shares 8–11 percentage points smaller, over an average of around 84%. Interestingly and consistently with our proposed mechanism, panels B and C show that for classes in the second year (panel B) and later (panel C), differences across ability groups fade out. In fact, class size in the second year of the program declines to an average of about 300 from 390 in the first year and further to 96 for other elective courses.

Notice that the differences in sample characteristics between the survey respondents and the experimental participants, mostly related to gender and high school grades, should not affect the comparability of results. Female students do not seem to be particularly different from their male classmates in terms of attendance behaviour and the differences in ability, which are presumably due to some degree of grade inflation, are mostly limited to the very upper part of the distribution.

All the courses considered in panels A and B of Table 8 are compulsory, so results should no be affected by selection issues. Nevertheless, in order to produce additional evidence on attendance patterns by ability and classroom scenarios, in the survey, we have also implemented a simple choice experiment. Students were presented with a picture of a very crowded and a rather empty classroom (in random order) and were asked to indicate whether they would be more likely to attend that class online or in person. Answers were given on a scale ranging from 1 to 7, with 1 corresponding to “*certainly online and not in person*” and 7 to “*certainly in person and not online*”. Table 9 reports the coefficients of the ability variables obtained from ordered logit regressions with the responses to the choice experiment as dependent variable and the same structure of explanatory variables as in Table 8.

Results clearly confirm that also when confronted with hypothetical scenarios, the high ability students are more likely to attend classes online than in person and especially so when the class is crowded. For example, based on the estimates produced

41. The set of controls includes gender, age, year of enrolment in the current program, language of the survey (students could choose to do it in French or English), exact time of survey start, time of survey completion, an indicator for whether the survey was completed (incomplete responses were still recorded), and number of usual hours of own study.

42. Average class size for first-year students in the academic year 2019–2020 was 390.

TABLE 8. Self-reported attendance.

	No controls	With controls ^a	No controls	With controls ^a
Panel A: First-year students, compulsory 1st-year classes				
High school grade ^b	0.007 (0.094)	−0.056 (0.099)	—	—
<i>By ability group^c</i>				
1 = Mid ability	—	—	0.032 (0.034)	0.018 (0.035)
1 = High ability	—	—	−0.084 (0.056)	−0.116** (0.057)
Obs.	164	162	164	162
<i>Descriptive statistics of the dep. variable:</i>				
Mean	0.841	0.843	0.841	0.843
Std. dev.	0.203	0.202	0.203	0.202
Panel B: Compulsory 2nd-year classes				
High school grade ^b	0.023 (0.087)	−0.054 (0.095)	—	—
<i>By ability group^c</i>				
1 = Mid ability	—	—	−0.032 (0.040)	−0.045 (0.041)
1 = High ability	—	—	−0.015 (0.057)	−0.066 (0.062)
Obs.	229	228	229	228
<i>Descriptive statistics of the dep. variable:</i>				
Mean	0.744	0.744	0.744	0.744
Std. dev.	0.243	0.244	0.243	0.244
Panel C: Other 2nd-year + classes				
High school grade ^b	−0.051 (0.095)	−0.113 (0.104)	—	—
<i>By ability group^c</i>				
1 = Mid ability	—	—	−0.065 (0.043)	−0.072 (0.044)
1 = High ability	—	—	−0.021 (0.062)	−0.037 (0.067)
Obs.	218	217	218	217
<i>Descriptive statistics of the dep. variable:</i>				
Mean	0.763	0.762	0.763	0.762
Std. dev	0.261	0.261	0.261	0.261

Notes: The dependent variable is the self-reported share of lectures attended in person in a normal term.

Robust standard errors in parentheses.

a. Controls include gender, age, year of enrolment in the current program (only in panels B and C), language of the survey (students could choose to do it in French or English), exact time of survey start, time of survey completion, an indicator for whether the survey was completed (incomplete responses were still recorded), and number of usual hours of own study.

b. Normalised high school leaving grade.

c. Low/mid/high based on percentiles of the distribution of high school grades. Low = 1st–20th, mid = 20th–80th, and high = 80th–100th. Low is the excluded category.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

TABLE 9. Attendance choice experiment.

	No controls		With controls ^a	
	Crowded ^b classes	Non-crowded ^c classes	Crowded ^b classes	Non-crowded ^c classes
High school grade ^d	− 1.205** (0.492)	− 0.468 (0.501)	− 1.116** (0.519)	− 0.478 (0.534)
Obs.	387	387	387	387
	Crowded ^b classes	Non-crowded ^c classes	Crowded ^b classes	Non-crowded ^c classes
1 = Mid ability ^e	− 0.062 (0.210)	− 0.074 (0.212)	0.019 (0.217)	− 0.132 (0.219)
1 = High ability ^e	− 0.599* (0.315)	− 0.427 (0.314)	− 0.535 (0.329)	− 0.455 (0.333)
Obs.	387	387	387	387

Notes: Ordered logit coefficients.

a. Controls include gender, age, year of enrolment in the current program, language of the survey responses, exact time of survey start, time of survey completion, an indicator for whether the survey was completed, and number of usual hours of own study.

b. The dependent variable is a scaled indicator about whether the respondent would attend an hypothetical crowded class in person or online (1 = certainly online; 7 = certainly in person).

c. The dependent variable is a scaled indicator about whether the respondent would attend an hypothetical non-crowded class in person or online (1 = certainly online; 7 = certainly in person).

d. Normalised high school leaving grade.

e. Low/mid/high based on percentiles of the distribution of high school grades. Low = 1st–20th, mid = 20th–80th, and high = 80th–100th. Low is the excluded category.

Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

with control variables, students in the high ability group are predicted to certainly attend in person (response category 7) a crowded class with a 10% probability against 16.5% for the students in the low ability group.

The results from our survey nicely complement the literature on students' attendance. Most papers in this area show that differences in attendance across the ability distribution are rather small and non-significant. For example, the meta-analysis covering over 70 published studies in Credé, Roch, and Kieszczynka (2010) confirms that class attendance displays "[...] *relative independence from SAT scores and high school GPA and weak relationship with student characteristics such as conscientiousness and motivation*" (page 273). Dolton et al. (2003) find that in their sample of students from the University of Malaga "*the correlation between pre-university scores and attendance is very low, at 0.11*" (page 550). Similarly, Arulampalam, Naylor, and Smith (2012) analyse classroom absences in a sample of second-year students at a UK university using first-year average grades as a measure of pre-determined ability.

Their summary statistics (Table 1, p. 367) show differences in grades between students missing few and many classes are very small and insignificant.⁴³

To the best of our knowledge, we are the first to produce evidence about attendance across types of classroom settings, suggesting that high-ability students might attend less in large crowded classes and more in smaller more interactive ones, eventually leading to no significant differences with other ability groups when averaging across all classes. This has important implications as it highlights that live-streaming might substitute different modes of attendance. One clear implication of this mechanism is that introducing live-streaming unambiguously increases welfare of all students: the most able ones enjoy higher grades while the least able, despite their lower grades, save on attendance costs.⁴⁴

The potential mechanism that we have put forward in this section seems consistent both with the results of our main analysis and also with the additional findings from our survey (and with the existing literature on attendance). The implications for the distribution of streaming take-up across ability groups depend on the exact form of the distribution of the shocks to the cost of attendance. In our data, take-up is rather low and homogeneous across ability groups, which would be consistent, for example and among other distributions, with rare but large (upward) variations to the cost of attendance.

5.1. *Alternative Mechanisms*

Of course, the interpretation of the results that we proposed in the previous section is not the only possible one, especially if students are not fully rational or uninformed about the costs and benefits of alternative learning modes.

For example, the negative effect of streaming on grades for the low-ability student might be explained by a lack of autonomy and self-control. By contrast, high-ability students might possess the motivation and ability to work autonomously that generate positive effects on grade when comparing with a crowded and noisy classroom. In order for this explanation to also rationalise the low take-up rates, it should be coupled with the assumption that students are myopic and/or over-/under-optimistic about the effects of the new technology on their learning. Otherwise, low-ability students should consistently use the streaming platform less often than high ability students, a behaviour that we do not observe in the experimental data.

Of course, the decision to attend in person might also be importantly affected by other considerations, such as the desire to socialise with friends, which may also potentially vary across ability groups. In fact, students of different abilities might use the streaming technology for different reasons. For example, the high-ability might

43. Also Romer (1993) presents evidence that, although indirectly, supports our model. He uses a sample of university students to run simple regression models with exam performance as a dependent variable and classroom attendance an explanatory factor and finds that adding prior GPA as a control does not change the coefficient on attendance significantly, suggesting a null or low correlation between attendance and prior GPA. Similar evidence is also presented in Bratti and Staffolani (2013) and Stanca (2006).

44. See also the discussion at the end of Online Appendix C.

stream when the quality of learning in the classroom is low, that is, when the class is crowded or when the class covers notions that they can learn easily on their own. Low-ability students might instead stream for reasons uncorrelated to the quality of learning in the classroom, such as bad (or good) weather, absence of their best friends, and so forth. Further research is certainly needed to improve our understanding of the way students interact with new teaching technologies.

6. Conclusions

In this paper, we present evidence from a randomized field experiment in which we offered access to live-streamed lectures to university students. Our results show that

1. students use the streaming service only rather rarely, about 10% of the times they have access;
2. offering the service has small effects on attendance in class, approximately 8 students out of 100 do not go to class; and
3. attending classes on the live streaming platform has positive effects on exam grades for high-ability students and negative effects for low-ability ones.

We discuss possible mechanisms that could rationalise these findings. One such mechanisms that seems consistent with the data postulates that students use the streaming service only when random events make the cost of attending in class particularly high. In the absence of such shocks, students of all ability levels prefer attending in class. In the counterfactual scenario in which streaming is not available and the shock hits, high-ability students prefer own study (no attendance), whereas low-ability ones still attend in class due to the higher value added of the professors' lectures. Hence, the heterogeneous effects of the experiment on grades emerge due to different counterfactuals: high-ability students substitute no attendance with streaming and low-ability students substitute in-class attendance with streaming.

This mechanism predicts higher attendance by low-ability students, at least in the large introductory classes with limited interactions with teachers that are in our experiment. We provide supporting evidence of this implication using data from an online survey of students.

One important element of innovation in our paper is the novel research design. We randomise access to the streaming service on a weekly basis and we map questions in the final exams to the weeks of term when the required material was covered. Hence, we can exploit variation both across and within students and weeks. This setup generates enough experimental variation to identify treatment effects across the distribution of students ability quite precisely.

Our results offer useful guidance for the design of education policies. Our evidence on take-up suggests that students have a general preference for classroom attendance, hence the use of distance learning technologies, at least in the form of simply streaming traditional classroom lectures, is unlikely to solve problems of

physical overcrowding. Of course, it is hard to say whether other technologies would have different effects but, at a minimum, our results suggest caution with the idea that distant learning tools can be used to reduce class size. Our theoretical discussion further highlights that the availability of the streaming service can also be seen as an education amenity that may increase the consumption value of college education, even if it comes at the cost of less effective learning for some students. Live-streaming services can thus be seen as an amenity offered by universities to attract students.

Finally, the heterogeneous effects on grades that we document in our experiment further indicate that online learning can potentially exacerbate education inequalities. One obvious policy implication would be offering such distance learning tools on a merit base, only to students at the upper end of the ability distribution. An alternative option is to offer the service to all students, highlighting that the effect on grades is beneficial only if it is used in occasion where attending is not possible or too costly.

Our analysis is specific to a very common learning environment, namely, large introductory classes with many students and rather standardised content. Hence, despite being a single case study it easily generalises to many similar settings. However, it is also fair to say that it is unlikely to generalise to very different contexts, especially those where the complementarity between the teachers' and the students' abilities are more important. In small classes, for example, student–teacher interactions may benefit the good students the most. Further studies in this direction would be important to explore the implications of distance learning technologies for higher education and ultimately for the design of education policies.

Appendix A: Additional Figures and Tables

Dear students,

In the context of the streaming project, you will have access to live streaming for all the courses you are attending, in the following weeks of the spring term :

Term week	Week of year	Streamed course available to you :
3	10	no
4	11	no
5	12	yes
6	13	yes
7	14	yes
8	15	no
9	16	Easter Holiday
10	17	yes
11	18	yes
12	19	yes
13	20	no
14	21	yes

To watch the live videos of the classes, you need to connect to the following web page :

<https://cms.unige.ch/gsem/streaming>

using your ISIs credentials (your usual login and password for UniGe services). Your access will be active 10 minutes before the beginning of the class. The first time you connect to this service, you might be asked to choose a teaching entity; select "Université de Genève" (as shown in the image):

FIGURE A.1. Notification email.



FIGURE A.2. Classroom pictures.

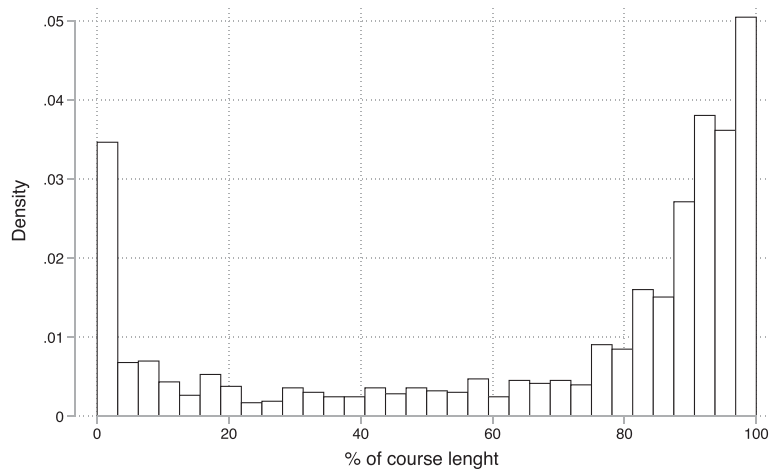


FIGURE A.3. Distribution of connection time.

TABLE A.1. Mean comparison test of variables by condition.

Variable	Treatment groups compared ^a	<i>p</i> -value
Swiss = 1	NA v SA	0.786
Swiss = 1	NA v AA	0.538
Swiss = 1	SA v AA	0.604
Female = 1	NA v SA	0.032
Female = 1	NA v AA	0.213
Female = 1	SA v AA	0.575
Age in years	NA v SA	0.027
Age in years	NA v AA	0.039
Age in years	SA v AA	0.574
High school grade ^b	NA v SA	0.596
high school grade ^b	NA v AA	0.961
high school grade ^b	SA v AA	0.547
Swiss high school	NA v SA	0.453
Swiss high school	NA v AA	0.714
Swiss high school	SA v AA	0.78
French high school	NA v SA	0.956
French high school	NA v AA	0.801
French high school	SA v AA	0.788
Mother with college	NA v SA	0.717
Mother with college	NA v AA	0.369
Mother with college	SA v AA	0.432
Father with college	NA v SA	0.953
Father with college	NA v AA	0.545
Father with college	SA v AA	0.472
Course-drop ^c	NA v SA	0.772
Course-drop ^c	NA v AA	0.586
Course-drop ^c	SA v AA	0.682
Grade ^d	NA v SA	0.3
Grade ^d	NA v AA	0.989
Grade ^d	SA v AA	0.302

Notes: a. NA: never access, SA: sometimes access, AA: always access.

b. Average normalised high school leaving grade.

c. Share of students not taking at least one exam.

d. Average grade normalised within each course.

TABLE A.2. Effect of the experiment on course-drop.

Endog. var:	ITT ^b	ATT ^b	
		Take-up ^c	Streaming ^d
All students	0.003 (0.002)	0.108 (0.084)	0.359 (0.286)
<i>By ability group^c</i>			
Low	0.008 (0.006)	0.287 (0.189)	0.920 (0.631)
Mid	0.002 (0.003)	0.068 (0.108)	0.238 (0.375)
High	0.001 (0.006)	0.043 (0.186)	0.134 (0.625)
Obs.	2545	2545	2545

Notes: Robust standard errors in parentheses.

The coefficients are produced by linear probability models where each observation is a student in a course (with student random effects). The dependent variable is a dummy equal to 1 if the student has a valid exam grade for the course and zero otherwise. The table reports the coefficients on the main regressors of interest, which varies by column. In column (1), it is the number of weeks the student had access to the streaming platform during the term. In column (2), it is a dummy equal to one if the student streamed the course at least once during term. In column (3), it is the number of weeks that student streamed during the term. In columns (2) and (3), the main regressors are instrumented with the number of weeks the student had access to the streaming platform during the term (the regressor of column (1)). In the lower panel, the same regressions are repeated with interactions of the main regressors with dummies for the three ability groups. All regressions include the following controls: course fixed effects, high school grade, a gender dummy, a dummy for Swiss nationals, dummies for enrollment cohorts (early, late, and regular enrollers), and nationality of the high school diploma (Swiss, French, and other nationalities).

a. take-up = 1 if student streamed the course at least once during term.

b. Streaming = number of weeks that student streamed the course over number of weeks with access.

c. Low-mid-high based on percentiles of the distribution of high school grades. Low = 1st–20th, mid = 20th–80th, and high = 80th–100th.

TABLE A.3. Probit estimation of weekly take-up.

Dep. var:	Take-up ^a
High school grade	1.405 (0.653)
(High school grade) ²	– 2.013 (0.803)
Obs.	14373
Students	1118

Notes: Robust standard errors in parentheses.

a. Estimation includes course and week fixed effects, a gender dummy, a dummy for Swiss nationals, dummies for enrollment cohorts (early, late, and regular enrollers), and nationality of the high school diploma (Swiss, French, and other nationalities). High school grade normalized based on scales and passing grades of country delivering the diploma.

TABLE A.4. First-stage estimates of ATT for grades.

Random effects ^b				Fixed effects ^b		
all students	0.123 (0.003)			0.118 (0.003)		
<i>By ability group^a</i>						
	Low	Mid	High	Low	Mid	High
Low	0.106 (0.003)	0.001 (0.005)	0.000 (0.003)	0.105 (0.003)	0.000 (0.005)	0.000 (0.003)
Mid	− 0.000 (0.002)	0.135 (0.003)	0.000 (0.002)	− 0.000 (0.002)	0.133 (0.003)	0.000 (0.002)
High	− 0.000 (0.003)	0.000 (0.006)	0.100 (0.003)	− 0.000 (0.003)	− 0.000 (0.006)	0.097 (0.003)
Obs. ^c	23766	23766	23766	23766	23766	23766

Notes: Robust standard errors in parentheses.

a. Low-mid-high based on percentiles of the distribution of high school grades. Low = 1st–20th, mid = 20th–80th, and high = 80th–100th. Columns report the first-stage estimates for each endogenous variable: low/mid/high are the interactions of low/mid/high-ability indicators with the streaming indicator. Rows are the excluded-instrument interactions of low/mid/high-ability indicators with the assignment indicator.

b. Estimation includes course and week fixed effects, a gender dummy, a dummy for Swiss nationals, dummies for enrollment cohorts (early, late, and regular enrollers), nationality of the high school diploma (Swiss, French, and other nationalities), and normalized high school grade for the pooled specification.

c. One observation per student-course-week.

TABLE A.5. Experimental effects on share of correct answers (robustness to different definitions of ability groups).

	Random effects			Fixed effects		
	ITT ^a	ATT ^b	OLS ^c	ITT ^a	ATT ^b	OLS ^c
<i>Percentiles of ability, baseline:^d</i>						
1st–20th	–0.019 (0.009)	–0.179 (0.098)	–0.020 (0.027)	–0.019 (0.009)	–0.178 (0.102)	–0.024 (0.027)
21st–80th	–0.000 (0.006)	–0.002 (0.047)	–0.003 (0.013)	0.000 (0.006)	0.001 (0.050)	–0.005 (0.013)
81th–100th	0.025 (0.010)	0.245 (0.111)	–0.011 (0.026)	0.023 (0.011)	0.241 (0.121)	–0.011 (0.028)
<i>Percentiles of ability:^d</i>						
1st–10th	–0.029 (0.013)	–0.219 (0.109)	–0.050 (0.032)	–0.032 (0.013)	–0.244 (0.114)	–0.058 (0.030)
11th–90th	0.001 (0.005)	0.012 (0.046)	0.003 (0.012)	0.002 (0.005)	0.014 (0.049)	0.001 (0.012)
91st–100th	0.026 (0.014)	0.246 (0.148)	–0.058 (0.040)	0.027 (0.015)	0.265 (0.158)	–0.061 (0.042)
<i>Percentiles of ability:^d</i>						
1st–30th	–0.018 (0.008)	–0.144 (0.067)	–0.025 (0.019)	–0.018 (0.008)	–0.147 (0.070)	–0.032 (0.019)
31th–70th	0.005 (0.007)	0.039 (0.057)	0.005 (0.015)	0.006 (0.007)	0.043 (0.060)	0.003 (0.016)
71st–100th	0.015 (0.009)	0.138 (0.086)	–0.004 (0.022)	0.015 (0.009)	0.143 (0.095)	–0.002 (0.023)
Obs. ^g :	23766	23766	23766	23766	23766	23766
<i>Descriptive statistics of the dependent variable:</i>						
Mean	0.545	0.545	0.545	0.545	0.545	0.545
Sd	0.378	0.378	0.378	0.378	0.378	0.378

Notes: Robust standard errors in parentheses. All regressions include the following controls: course fixed effects, week-wave fixed effects, a gender dummy, a dummy for Swiss nationals, dummies for enrollment cohorts (early, late, and regular enrollers), nationality of the high school diploma (Swiss, French, and other nationalities), and normalized high school leaving grade (only in the panel *all students*).

a. Columns named ITT show OLS estimates of the share of correct answers on assignment status (=1 if student has access in the given week for the given course).

b. Columns named ATT show IV estimates of the share of correct answers using assignment status (=1 if student has access in the given week for the given course) as instrument for the streaming indicator (=1 if student connected to streaming in the given week for the given course).

c. Columns named OLS show OLS estimates of the share of correct answers on streaming take-up (=1 if student connected to streaming in the given week for the given course).

d. Ability groups based on percentiles of the distribution of high school grades. The reported estimates are the coefficients on the interactions of the ability group indicators with the assignment status (column ITT), the instrumented streaming indicator (column ATT) or the streaming indicator (column OLS).

Appendix B: Theory

In this section, we develop a simple model of learning and attendance choice that rationalizes our empirical results and highlights plausible mechanisms. Our objective is to provide a simple setup that is able to rationalize our two main empirical findings, namely, (i) the heterogeneity of the treatment effect on exam grades by ability and (ii) the low and homogeneous take-up rates by ability.

Taken together these two results might appear counter-intuitive: If streaming is better than attending for high-ability students why are they not streaming more often? Symmetrically, if streaming has negative effects on low-ability students why are they not streaming less than other students? We propose an explanation based on two key assumptions. First, we postulate that the cost of in-class attendance is affected by random shocks, such as weather conditions or sickness. Second, we assume that the benefit of in-class participation are larger for low ability students than high ability ones. The combination of these two hypotheses generates different counterfactuals to the use of streaming service for students of different abilities: in the absence of streaming, low ability students still attend in class whereas the high ability study on their own. When streaming is available, all students use it when hit by attendance shocks. The theory we build on these assumptions generates homogeneous take-up and heterogeneous treatment effects by ability, as in our data.

Our model builds on the existing literature on the education production function (Hanushek 1986) but departs from the standard model in two notable ways. First, we introduce a menu of different learning technologies from which students choose optimally. Second, we allow differences in ability to affect students' returns to effort of the different technologies and, for simplicity, we assume that the utility cost of effort is the same for all students regardless of their ability.

Environment. Consider a student who needs to choose a specific learning technology from a menu of available alternatives and exerts costly effort to study, learn and obtain a grade in academic tests or exams.⁴⁵ Students value high academic performance as measured by exam grades.⁴⁶

Students choose the learning technology and the level of effort in order to maximize the following utility function:

$$w_i(x_i, e_i) = x_i - \frac{e_i^2}{2} - c^j, \quad (\text{B.1})$$

where i indexes the student and j the learning mode. x_i is academic performance (e.g., grades), e_i denotes effort and c^j is the utility cost of adopting learning mode j .

Academic performance is the output of effort and ability, processed through the chosen learning technology. A menu of three technologies is potentially available to

45. In what follows, we use the terms *learning technology* and *attendance* or *learning mode* interchangeably.

46. We do not enter the debate on whether exam grades correctly measure learning.

the students:

$$\text{in class attendance: } x_i = x^a(e_i) = \beta_i^{\gamma^a} e_i, \quad (\text{B.2})$$

$$\text{live streaming: } x_i = x^s(e_i) = \beta_i^{\gamma^s} e_i, \quad (\text{B.3})$$

$$\text{own study or no attendance: } x_i = x^n(e_i) = \beta_i^{\gamma^n} e_i, \quad (\text{B.4})$$

where $\beta_i \in (0, 1)$ represents the ability level of student i and $\gamma^j \in (0, 1/2]$ is a technological parameter describing how ability affects the returns to effort.⁴⁷ This parametrization implies that, given the production technology, high-ability students have higher returns to effort compared to their low-ability peers.

Assumptions. We further impose the following set of assumptions.

ASSUMPTION A1. Classroom attendance is the most efficient learning technology:

$$\gamma^n > \gamma^s \geq \gamma^a.$$

ASSUMPTION A2. The costs of both no attendance and streaming are normalized to zero:

$$c^s = c^n = 0.$$

ASSUMPTION A3. The cost of classroom attendance is random. In normal times it is normalized to zero and with probability p it is hit by a shock of random intensity (e.g., traffic congestion due to weather conditions or sickness of varying intensity):

$$c^a = \begin{cases} u > 0 & \text{with probability } p \\ 0 & \text{with probability } 1 - p. \end{cases}$$

with u uniformly distributed over $[\underline{u}, \bar{\beta}u]$. The shock u is realized before the choice of attendance mode is made.

ASSUMPTION A4. Only students above the minimum level of ability $\underline{\beta}$ enrol at university and hence appear in our data:

$$\beta_i > \underline{\beta}.$$

Assumption (A1) implies that, while high-ability students have higher returns to effort compared to their low-ability peers, the low-ability students benefit relatively more from a more efficient learning technology. We believe that this is a valid assumption in the setting that we study, where the large class sizes do not allow intense interactions between students and teachers. High-ability students can read the material

47. We interpret β as parameterizing “ability” of students as we have a credible proxy of ability in the empirical part of the paper. However, it is also possible to theoretically interpret this parameter as being an innate “motivation” for study. Arguably both ability and motivation correlate strongly and both interpretation lead to the same implications.

on a textbook or lecture notes and understand most of it, hence the value added of the professor's explanations is limited. Low-ability students, instead, really benefit from such explanations and would only understand a smaller part of the material if they were studying on their own. In smaller classes, where interactions are more intense, the complementarity between student and teacher abilities would be much more important and potentially make the returns to attendance much higher for the good students. (A1) is consistent with the evidence we produce from our online survey (see Section 5 and Online Appendix D) and also with a number of papers highlighting the benefits of attending classes (Romer 1993; Dobkin, Gil, and Marion 2010; Arulampalam, Naylor, and Smith 2012; Bratti and Staffolani 2013; Kassarnig et al. 2017).

(A2) is a simplifying assumption limiting the number of cases to be analysed. Introducing a positive cost of streaming (and no attendance) would modify the decisions of the very low and very high-ability students, which would multiply the number of outcomes without changing qualitatively our results.

(A3) encapsulates the idea that attendance is overall costly for students (opportunity cost of time and transportation costs) and is the reason why students sometimes do not attend or stream despite higher return from class attendance. The normalization of c^a to zero in the absence of shocks is a simplifying assumption, relaxing it would complicate the analysis without changing its main implications. We could also allow the cost of effort to be larger for less able students, as in some well-known models of education, but this would be an unnecessary complication in our setting. More generally, our model can be extended in various dimension (i.e., by adding behavioural assumptions such as self-control issues or myopic evaluation of return to efforts) but we choose to keep it simple to highlight what we believe is the most salient mechanism.

(A4) is also a simplifying assumption limiting the number of cases to study and clarifying the exposition.

Optimal Choice. We assume that students maximize utility by choosing their mode of attendance and their level of effort. Optimal effort conditional on learning technology j is:

$$e_i^j = \beta_i^{\gamma^j}. \quad (\text{B.5})$$

Consequently, optimal grades are $x_i^j = \beta_i^{2\gamma^j}$ and the mode of attendance is chosen based on comparing indirect utility under the three alternatives:

$$w_i^j = \frac{\beta_i^{2\gamma^j}}{2} - c^j. \quad (\text{B.6})$$

Figures B.1–B.3 illustrate the different returns of ability to efforts for the three technologies (equation B.5). It is noteworthy that the optimal level of effort (and the associated welfare) increases with β_i , the student's ability. However, the figures make clear that low-ability students benefit relatively more from a more efficient learning technology than their high-ability peers: fixing β , we observe a higher increase in optimal efforts when moving toward more efficient technology—from B.1 to B.3—for

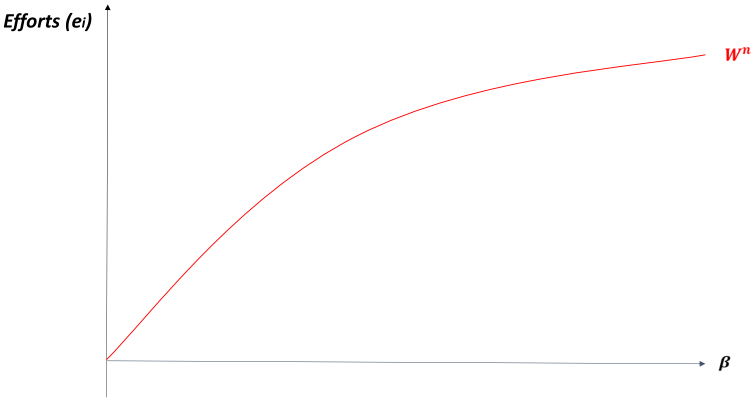


FIGURE B.1. No-attendance.

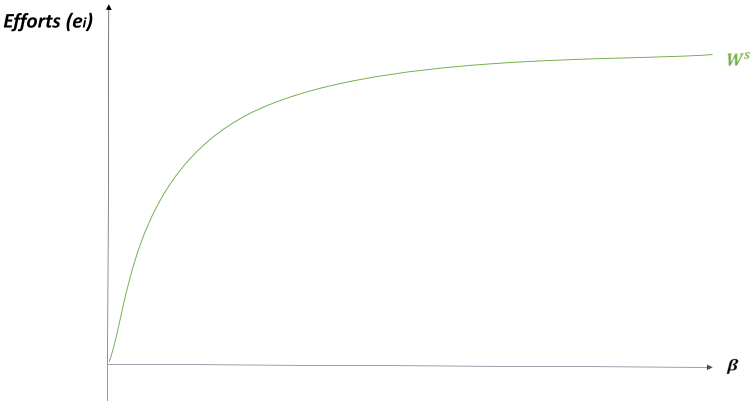


FIGURE B.2. Streaming.

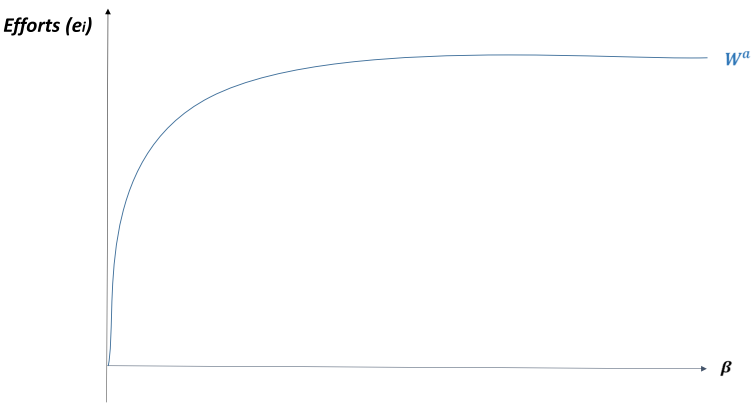


FIGURE B.3. Attendance.

low ability than high ability (implied by the different degree of concavity of grades with respect to β). This illustrates the intuitive idea that students at the upper end of the ability distribution will comprehend most material even without the presence of a professor (implying a low difference in returns between technologies); whereas this input is comparatively of greater importance for students at the lower end of the ability distribution.

Solution Without Streaming. In the context of this model, our experiment can be seen as a manipulation of the menu of available learning technologies. When access to the streaming platform is offered, all three technologies are available, otherwise students can only choose between in-class attendance and own study. Let us first consider the scenario in which streaming is not available. First, notice that in the absence of shocks students always go to the classroom: when $c^a = 0$, $w_i^a > w_i^n$. When facing attendance shocks, students switch to non attendance only if the shock is large enough, namely, when

$$u > \frac{\beta_i^{2\gamma^a} - \beta_i^{2\gamma^n}}{2}. \quad (\text{B.7})$$

Importantly, this threshold varies (negatively) with ability. Hence for every realization of the shock u , we can define a critical type $\bar{\beta}_u$ such that students with ability above $\bar{\beta}_u$ would switch to non-attendance, whereas those below would still go to class.⁴⁸

Figure B.4 illustrates the functioning of the model and plots indirect utility by ability and attendance modes, with and without shocks. For clarity of exposition, we have set γ^n to 1/2 so that indirect utility when not attending is linear in ability. This is merely a simplifying assumption, the implications of the model hold also with concave returns to non-attendance as what matters for our results is the ordering implied by [A1]. The figure illustrates that only the high-ability students switch from attending to not attending when exposed to exogenous shocks. Equation (B.7) further shows that, as the magnitude of the shock increases, the threshold $\bar{\beta}_u$ decreases and more students stop going to class. Given the assumption that class attendance is the most effective learning technology (Assumption A1), this switch generates a decrease in academic performance.

Solution with Streaming. We now add streaming to the menu of available learning technologies. Given Assumption A1, it is evident that in the absence of shocks all students attend in class and the take-up rate of the streaming service is zero. As we will show in Section 4, this parametrization is consistent with our data, where almost no student uses the streaming platform every time she has access to it (see Figure 4).

Figure B.5 shows what happens with shocks and rationalizes one of the most salient results of our analysis, namely, the positive effect of the streaming service on the grades of high-ability students and the negative effect on those of the low-ability ones. To simplify the exposition we only consider shocks that are large enough to induce streaming for students of all abilities (conditional on the participation

48. $\bar{\beta}_u$ is defined as $(\bar{\beta}^{2\gamma^a} - \bar{\beta}^{2\gamma^n})/2 = u$.

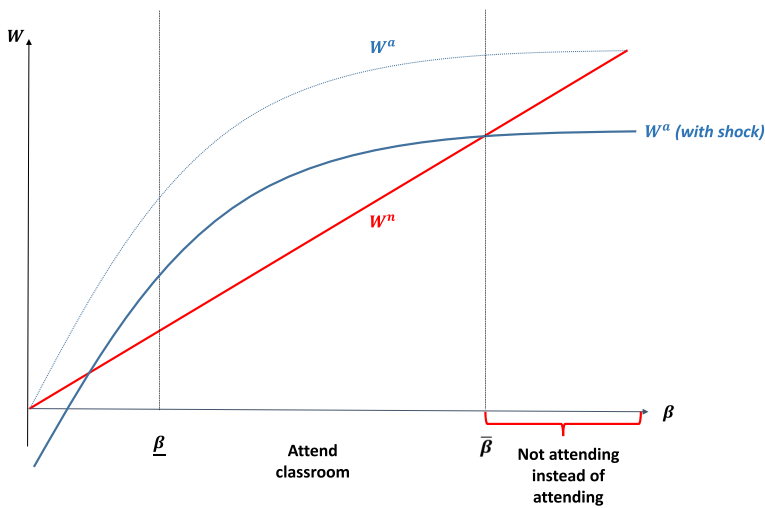


FIGURE B.4. Indirect utility and shocks without streaming.

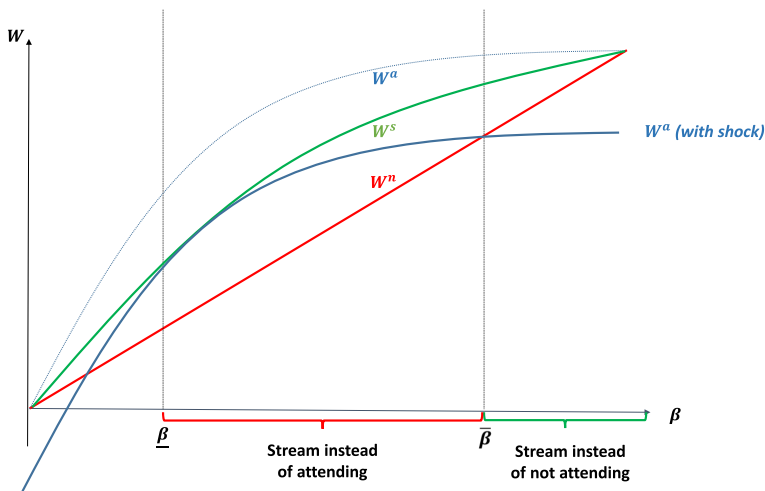


FIGURE B.5. Indirect utility and shocks with streaming.

constraint $\beta_i > \underline{\beta}$.⁴⁹ When hit by a shock, the high-ability students ($\beta_i > \bar{\beta}_u$) choose to stream instead of not attending, thus inducing a positive effect on grades, whose magnitude increases with the student's ability.

Students at the lower end of the ability distribution ($\beta_i < \bar{\beta}_u$) also use the streaming service but in this case they substitute away from in class attendance,

49. With smaller shocks the effect on grades would be null for some students at the bottom of the ability distribution.

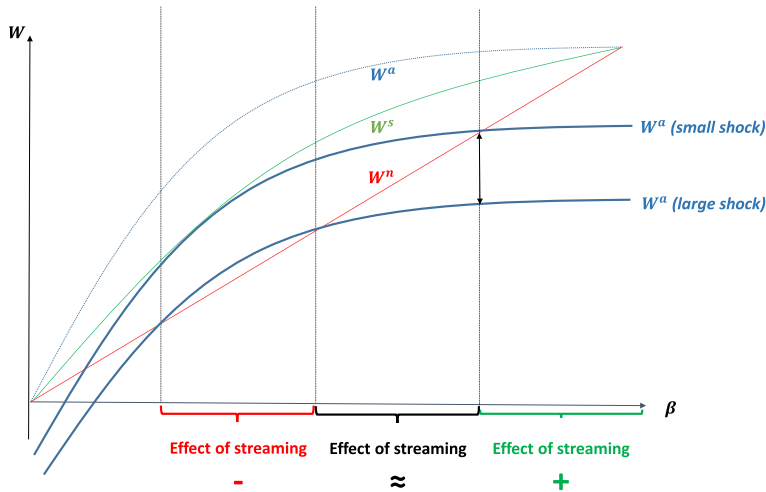


FIGURE B.6. The effect of streaming on grades.

the most efficient learning technology. Hence, the effect of streaming on grades is negative. Contrary to the positive effect for high-ability students, the magnitude of this negative effect decreases with ability.

Figure B.6 highlights the differential effects of access to streaming on students of different abilities. Low-ability students substitute streaming to attendance and experience a drop in grades, which decreases with ability. High-ability students substitute streaming to no-attendance and experience an increase in grades, which is larger for higher ability students. The effect of streaming on students in the middle range of the ability distribution depends crucially on the size of the shock: they substitute streaming to attendance for small shocks and streaming to no-attendance for larger shocks. Hence, taking averages over different realizations of the shock and across a group of students in the middle of the distribution is likely to result in an effect on grades pretty close to zero, which is what we documented empirically in Section 4.

Summary of Results. This simple theoretical framework highlights that without access to streaming, most students attend class unless exogenous shocks make their cost of attendance too high. When these shocks happen, the low-ability students still attend classes, whereas the high-ability students tend to stay at home and study on their own. These high-ability students can read the material in the book and understand it easily even without the professors presenting and explaining it. In the terminology of the model, the return to effort in class and in own study are comparable for high-ability students. When streaming is available, most students use it when they are hit by (sufficiently large) attendance shocks. The least able students stream instead of going to class, thus experiencing lower grades, whereas the most able students stream instead of not attending, hence experiencing higher grades. For students in the middle range of the ability distribution, the prediction is ambiguous: they substitute streaming

TABLE C.1. Survey structure.

Section	Questions
1. About yourself	Age, gender, academic program, and year of enrolment
2. High school	Leaving grade and minimum and maximum passing grade
3. Self-reported attendance	Percentage of lectures attended in person in a normal semester, broken down by compulsory first-year classes, other compulsory classes and elective classes
4. Preferences for distance learning and own study	General preference for attending lectures online or in person (scale 1–7), number of hours of own study per week
5. Choice experiment of classroom attendance	Likelihood (scale 1–7) of attending classes online or in person when presented with picture of a rather crowded or a rather empty classroom. Pictures presented in random order
6. Choice experiment of classroom attendance and weather conditions	Likelihood (scale 1 to 7) of attending classes online or in person when presented with picture of a rather crowded or a rather empty classroom combined with pictures of a sunny or rainy day. Only two random combinations presented to each respondent.
7. Commuting	Zip code, country of residence, and commuting modes
8. Familiarity with distance learning tools	Familiarity (on a scale 1–7) with popular distance learning tools (e.g., Coursera)
9. Final comments	Open comments

to attendance for small shocks and streaming to no-attendance for larger shocks. Hence, taking averages across a group of students in the middle of the distribution is likely to result in an effect on grades pretty close to zero.

An important implication of the model is that introducing streaming unambiguously increases welfare for all students (as shown in Figure B.5): they are all better off, in terms of utility, when using streaming if they are hit by a large shock. The best students are better-off because of higher grades (at the cost of a slightly higher streaming cost with respect to not attending) while low ability students are better-off because they save on the high attendance cost, even if their grades are lower.

Appendix C: Online Survey on Attendance Behaviour

In order to investigate attendance patterns for different students and classroom settings, we implemented an online survey of all students enrolled in any program offered by the Faculty of Economics and Management at the University of Geneva. This includes 1 bachelor and 6 master programs, for a total of 1,209 students regularly enrolled in the academic year 2019–2020. Students were invited to participate in the survey by email and were sent up to six reminders between mid April and mid May 2020. The first recorded response dates 8 April 2020 and the last one 13 May 2020.

The survey questionnaire consisted of nine sections for a total of 27 questions, which students took on average about 8 minutes to complete. The survey could be answered in either French or English. Table C.1 synthetically describes the structure of the survey.

TABLE C.2. Descriptive statistics.

	First-year students		All students	
	Mean	Std. dev.	Mean	Std. dev.
Age (years)	20.26	2.56	21.29	2.35
1 = female	0.60	0.49	0.54	0.50
1 = first year	1.00	0.00	0.41	0.49
High school grade ^a	0.44	0.22	0.45	0.21
1 = survey in French ^b	0.858	0.350	.875	0.330
Time to complete (sec.)	386.70	556.96	434.04	567.08
1 = completed survey ^c	0.85	0.36	0.84	0.37
Invited students	398		922	
Total responses	186		455	
Valid responses ^d	183		450	

Notes: a. Standardised across school systems. See text for details.

b. The survey could be answered in English or French.

c. Answers of uncompleted surveys were recorded.

d. Due to inconsistent answers or missing values, some responses could not be used for the analysis.

We ask detailed information on high school grades in Section 2 because students come from a variety of different school systems and we need to compute a measure of pre-determined ability based on high school grades that can be compared across systems. Hence, we ask both the students grade and the minimum and maximum passing grades in their school systems. Then, we measure ability as (own grade – minimum passing grade)/(maximum-minimum passing grade), which results in a number between 0 and 1 that indicates the position of the student's grade in the interval between the minimum and the maximum grade that would allow obtaining the degree in his/her system. This is the same measure of ability that we use in our main experiment and throughout the main text.

Table C.2 reports some basic descriptive statistics, separately for first-year students, the most comparable to the students in our experiment, and all students. The survey was administered during the Coronavirus epidemics when the buildings of the University of Geneva were not accessible to students and most lectures were taking place online. This has probably contributed to increase the response rate, both because students were spending a lot of time in front of their computers and also because the issue of distance learning was clearly very salient. We did, however, take care and emphasising that, for many questions, students were supposed to answer with reference to a normal period of the academic year.

References

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

- Arulampalam, Wiji, Robin A. Naylor, and Jeremy Smith (2012). "Am I Missing Something? The Effects of Absence from Class on Student Performance." *Economics of Education Review*, 31, 363–375.
- Banerjee, Abhijit V. and Esther Duflo (2014). "(Dis)organization and Success in an Economics MOOC." *American Economic Review*, 104(5), 514–518.
- Bettinger, Eric P., Lindsay Fox, Susanna Loeb, and Eric S. Taylor (2017). "Virtual Classrooms: How Online College Courses Affect Student Success." *American Economic Review*, 107(9), 2855–2875.
- Bowen, William G., Matthew M. Chingos, Kelly A. Lack, and Thomas I. Nygren (2014). "Interactive Learning Online at Public Universities: Evidence from a Six-Campus Randomized Trial." *Journal of Policy Analysis and Management*, 33, 94–111.
- Bratti, Massimiliano and Stefano Staffolani (2013). "Student Time Allocation and Educational Production Functions." *Annals of Economics and Statistics*, 111/112, 103–140.
- Brown, Byron W. and Carl E. Liedholm (2002). "Can Web Courses Replace the Classroom in Principles of Microeconomics?" *American Economic Review*, 92(2), 444–448.
- Coates, Dennis, Brad R. Humphreys, John Kane, and Michelle A. Vachris (2004). "'No Significant Distance' Between Face-to-Face and Online Instruction: Evidence from Principles of Economics." *Economics of Education Review*, 23, 533–546.
- Credé, Marcus, Sylvia G. Roch, and Urszula M. Kieszczynka (2010). "Class Attendance in College: A Meta-Analytic Review of the Relationship of Class Attendance With Grades and Student Characteristics." *Review of Educational Research*, 80, 272–295.
- Deming, David J., Claudia Goldin, and Lawrence Katz (2012). "The For-Profit Postsecondary School Sector: Nimble Critters or Agile Predators?" *Journal of Economic Perspectives*, 26(1), 139–164.
- Deming, David J., Claudia Goldin, and Lawrence Katz (2013). "For-Profit Colleges." *The Future of Children*, 23, 137–163.
- Deming, David J., Claudia Goldin, Lawrence F. Katz, and Noam Yuchtman (2015). "Can Online Learning Bend the Higher Education Cost Curve?" *American Economic Review*, 105(5), 496–501.
- Deming, David J., Noam Yuchtman, Amira Abulafi, Claudia Goldin, and Lawrence F. Katz (2016). "The Value of Postsecondary Credentials in the Labor Market: An Experimental Study." *American Economic Review*, 106(3), 778–806.
- Dobkin, Carlos, Ricard Gil, and Justin Marion (2010). "Skipping Class in College and Exam Performance: Evidence from a Regression Discontinuity Classroom Experiment." *Economics of Education Review*, 29, 566–575.
- Dolton, Peter, Oscar D. Marcenaro, and Lucia Navarro (2003). "The Effective Use of Student Time: a Stochastic Frontier Production Function Case Study." *Economics of Education Review*, 22, 547–560.
- Escueta, Maya, Vincent Quan, Andre Joshua Nickow, and Philip Oreopoulos (2017). "Education Technology: An Evidence-based Review." NBER Working Paper No. 23744.
- Figlio, David, Mark Rush, and Lu Yin (2013). "Is It Live or Is It Internet? Experimental Estimates of the Effects of Online Instruction on Student Learning." *Journal of Labor Economics*, 31, 763–784.
- Hanushek, Eric A. (1986). "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature*, 24, 1141–1177.
- Jacob, Brian, Brian McCall, and Kevin Stange (2018). "College as Country Club: Do Colleges Cater to Students' Preferences for Consumption?" *Journal of Labor Economics*, 36, 309–348.
- Joyce, Ted, Sean Crockett, David A. Jaeger, Onur Altindag, and Stephen D. O'Connell (2015). "Does Classroom Time Matter?" *Economics of Education Review*, 46, 64–77.
- Kassarnig, Valentin, Andreas Bjerre-Nielsen, Enys Mones, Sune Lehmann, and David Dreyer Lassen (2017). "Class Attendance, Peer Similarity, and Academic Performance in a Large Field Study." *PLOS One*, 12, 1–15.
- McPherson, Michael S. and Lawrence S. Bacow (2015). "Online Higher Education: Beyond the Hype Cycle." *Journal of Economic Perspectives*, 29(4), 135–154.
- Oreopoulos, Philip and Uros Petronijevic (2018). "Student Coaching: How Far Can Technology Go?" *Journal of Human Resources*, 53, 299–329.

- Romer, David (1993). "Do Students Go to Class? Should They?" *Journal of Economic Perspectives*, 7(3), 167–174.
- Setren, Elizabeth, Kyle Greenberg, Oliver Moore, and Michael Yankovich (2020). "Effects of the Flipped Classroom: Evidence from a Randomized Trial." *Education Finance and Policy*, o:ja, 1–54.
- Stanca, Luca (2006). "The Effects of Attendance on Academic Performance: Panel Data Evidence for Introductory Microeconomics." *The Journal of Economic Education*, 37, 251–266.
- Swoboda, Aaron and Lauren Feiler (2016). "Measuring the Effect of Blended Learning: Evidence from a Selective Liberal Arts College." *American Economic Review*, 106(5), 368–372.
- U.S. Department of Education (2010). "Evaluation of Evidence-Based Practices in Online Learning: a Meta-Analysis and Review of Online Learning Studies." Technical Report, US Department of Education, Washington, D.C.
- Xu, Di and Shanna Smith Jaggars (2011). "The Effectiveness of Distance Education across Virginia's Community Colleges: Evidence from Introductory College-Level Math and English Courses." *Educational Evaluation and Policy Analysis*, 33, 360–377.
- Xu, Di and Shanna Smith Jaggars (2013). "The Impact of Online Learning on Students' Course Outcomes: Evidence from a Large Community and Technical College System." *Economics of Education Review*, 37, 46–57.

Supplementary Data

Supplementary data are available at [JEEA](https://academic.oup.com/jeea/article/19/4/2322/6067382) online.