

DISCUSSION PAPER SERIES

IZA DP No. 12298

**Distance Learning in Higher Education:
Evidence from a Randomized Experiment**

M. Paula Cacault
Christian Hildebrand
Jérémy Laurent-Lucchetti
Michele Pellizzari

APRIL 2019

DISCUSSION PAPER SERIES

IZA DP No. 12298

Distance Learning in Higher Education: Evidence from a Randomized Experiment

M. Paula Cacault

University of Geneva

Christian Hildebrand

University of St. Gallen

Jérémy Laurent-Lucchetti

University of Geneva

Michele Pellizzari

University of Geneva, CEPR and IZA

APRIL 2019

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Distance Learning in Higher Education: Evidence from a Randomized Experiment*

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) students use the live streaming technology only punctually, apparently when random events make attending in class too costly; (ii) attending lectures via live streaming lowers achievement for low-ability students and increases achievement for high-ability ones and (iii) offering live streaming reduces in-class attendance only mildly. These findings have important implications for the design of education policies.

JEL Classification: I20, I21, I23, I26

Keywords: EduTech, distance learning, live streaming

Corresponding author:

Michele Pellizzari
Institute of Economics and Econometrics
Geneva School of Economics and Management
University of Geneva
40 Bd du Pont d'Arve
CH-1211, Geneva 4
Switzerland

E-mail: michele.pellizzari@unige.ch

* This project has only been possible thanks to the support of many persons in the administration of the University of Geneva. We would like to thank Stéphane Berthet, Marcelo Olarreaga, Emmanuel Rousseaux, José Stultschnik, Maria-Pia Victoria-Feser, 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, Lammert Vrieling. Our gratitude also goes to all the students in the experimental courses. 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, Paris School of Economics. All errors remain our sole responsibility.

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 an ever-evolving offer of e-learning platforms, distance learning tools, MOOCs, and the like. Given the enormous increase in the demand for higher education around the world and the importance of education as a driver of growth, distance learning tools for colleges are attracting particular attention in this context (Woessmann, 2016). This class of technologies, especially in the form of distance broadcasting of lectures, can lower the marginal cost of university education and thus allow offering it to larger masses around the world (Deming, Goldin, Katz, and Yuchtman, 2015). In fact, the promise of cost savings fueled the very rapid emergence of an entire new sector of private for-profit colleges offering mostly on-line education (Bettinger, Fox, Loeb, and Taylor, 2017; Deming, Yuchtman, Abulafi, Goldin, and Katz, 2016; Koedel, Darolia, Martorell, Wilson, and Perez-Arce, 2014).

In this paper we provide experimental evidence on the impact of distance learning technology on student performance. We design and implement a randomized experiment in which first year bachelor students in a public university are randomly offered access to a live streaming platform for many of their compulsory courses.¹ Access to the platform is 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. Classroom attendance is always available to all students, allowing us to study students' choices of the attendance mode. Furthermore, we construct a mapping of all questions in the final exams to the weeks of the term in which the material was covered. Thus, we can exploit variation both across and within students and weeks to identify the effect of the technology on achievement.

Our analysis shows that (i) students of all abilities use distant live streaming only punctually, apparently when attending class is too costly (i.e. during epidemic outbreaks) and (ii) attending lectures via live streaming 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 effect) reduces the probability of answering correctly to exam questions by 2 percentage points for students in the lowest quintile of the ability distribution over an average probability of 55%.² The effect at the opposite end of the ability distribution (i.e. students in the top 20%) is positive and equal to 2.5 percentage points. Additionally, we find that offering access to the live streaming platform has only a modest effect on class size: for each 100 students with access there will be 8 fewer in class.

We rationalize these findings with a simple theoretical framework where students optimally choose a mode of attendance out of three alternatives: in-class, own study (i.e. no attendance) or online streaming, if available. Each of these modes is characterized by different utility costs and by different returns to study effort. We show that our empirical findings are consistent with a parametrization of the model where students only use the streaming service when they are hit by an attendance shock (e.g. illness, bad weather, etc.). A crucial result is that the counterfactual to streaming differ by ability. For good students the return to effort of own studying is only mildly lower than that of in-class attendance, hence they do not go to class when hit by the attendance shock. For the least able students, own study is very inefficient and

¹Live streaming is arguably the most serious contender to traditional education, it combines salient advantages of other e-learning technologies - i.e. a cost-efficient way of teaching to a massive number of students - while maintaining crucial features of traditional classroom attendance, such as the organized structure of a fixed classroom schedule.

²We measure ability using high school grades.

they tend to go to class even when hit by the shock. Hence, the counterfactual to streaming is no attendance for the good students and regular in-class attendance for the least able students, rationalizing the heterogeneous returns that we document in our empirical analysis. This set of results paints an overall consistent picture: students have a general preference for in-class attendance and use streaming only when random events make it too costly to attend; streaming has positive effects on learning if used as a substitute to no attendance and negative effects if it substitutes in-class attendance.

We believe that our study provides a notable contribution to the growing literature on *Edutech* and on distance learning in particular (Escueta, Quan, Nickow, and Oreopoulos, 2017; U.S. Department of Education, 2010). Despite the enthusiasm often surrounding these innovations, credible evidence about their impact on learning outcomes is still limited and the few 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) shows worse labor market outcomes for graduates of online for-profit colleges. Brown and Liedholm (2002) show suggestive evidence of substantially lower achievement for college students attending classes in mixed or fully online modes.

While there exist a myriad of descriptive studies (see U.S. Department of Education (2010) for a review), we are aware of only a handful of papers providing plausibly causal estimates of the use of distance learning tools in higher education. Three of them use instrumental variables, namely Bettinger et al. (2017), Coates, Humphreys, Kane, and Vachris (2004) and Xu and Jaggars (2013). All these papers use 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.³

A few other studies provide (relatively small-scale) experimental evidence. Alpert, Couch, and Harmon (2016), Figlio, Rush, and Yin (2013) and Joyce, Crockett, Jaeger, Altindag, and O'Connell (2015) have rather small samples of 300 to 700 students in an introductory microeconomics class and find mixed results. Bowen, Chingos, Lack, and Nygren (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 In Alpert et al. (2016) and Bowen et al. (2014) the experimental comparison is actually not between traditional in-person and distance teaching but rather between alternative combinations of the two, thus making it difficult to attach a clear interpretation to the findings.

Our study provides a novel combination of 3 important features. First, we exploit experimental variation over a relatively large sample of students, across multiple courses and with clear counterfactuals. In addition, we exploit variation across weeks of the term and exam questions for the same student and we eventually rely on over 23'000 student-course-week observations for our main results. Second, our setting allows us to investigate take-up decisions, something that could not be done in any of the previous papers due to the combination of small samples and participation incentives. Third, we focus on a relatively standard institution, public and not-for profit, a setting that resembles very much that of most higher education institutions around the world. As such, our results are more likely to generalize to many other settings.

³The 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).

⁴In Alpert et al. (2016) the incentive was a 5-point bonus on the course average grade.

The paper is organized as follows. In Section 2 we describe the institutional setting and the experiment. Section 3 presents a simple theoretical framework that is useful to guide the empirical exercise and interpret the findings. Section 4 describes the data that we use for the empirical analysis, which we introduce and discuss in Section 5. Section 6 concludes.

2 Institutional and experimental setup

2.1 General context

As part of a general effort towards 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 experiment it on a smaller scale. The authors of this paper were asked to design and implement such an experiment.

We decided to target classes in the bachelor program in economics and management, 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 3 semesters. All the compulsory courses of the first part are taught in (at least) 2 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.⁵The recordings are only available 24 to 48 hours after the lectures take place and this is problematic for the students for at least two reasons. First, their class schedules - especially in their first year - are extremely tight, hence it is difficult for them to find a slot of 2 hours to watch the recorded lectures. In most cases the only available times would be either early in the morning, late in the evening or in the weekends. Second, TA sessions requiring knowledge of the material presented in the lectures might take place before the 24-48 hours delay for the recordings to be available. In addition, the recording technology shows the video capture from the classroom computer (i.e. slides or electronic board) and the audio, hence any body-language is not visible. For all these reasons, students tend to use the recorded lectures mostly as reference when they revise the material and only very rarely as a substitute of in-class attendance.

2.2 Experiment setup

During two academic semesters, Spring 2017 (Feb-May) and Fall 2017 (Sep-Dec), we live-streamed all the lectures of 8 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

⁵The possibility to record lectures for later viewing was already available on a voluntary basis to all professors (provided their lectures took place in classrooms equipped with the necessary technology) since a few years at the University of Geneva.

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.

Table 1: Courses participating in the experiment

Term	Course	Language	Size	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

^a EM: Economics and Management. IR: International Relations.

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 Economics and Management (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.

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, i.e. covered exactly the same content, used the same problem sets, 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 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

⁶These 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.

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

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

⁹These are students in the first cohort, i.e. those who entered the university in September 2016, and who retook

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 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.

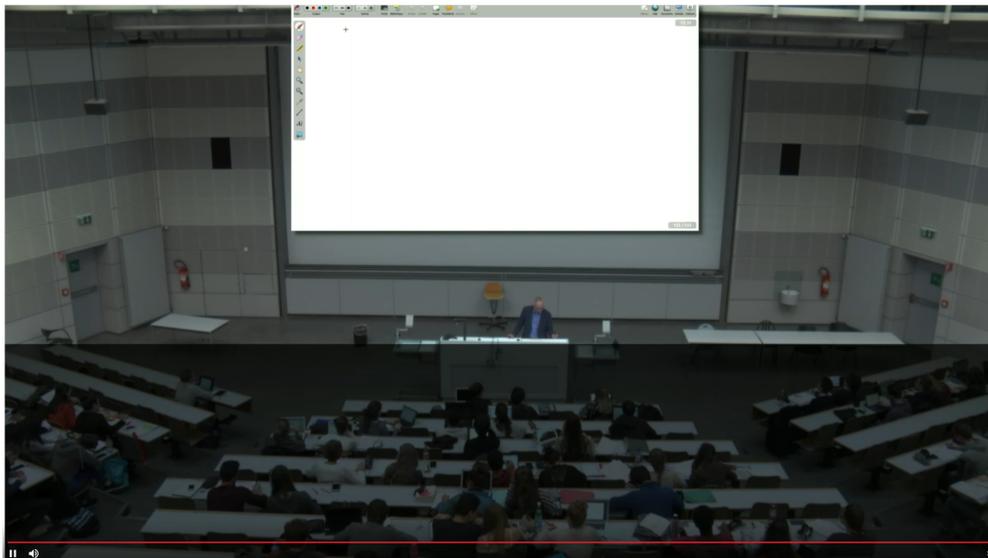


Figure 1: Live streaming - example

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

At the University of Geneva terms consist of 13 weeks (plus 1 week of mid-term break). In week one, the professors of the participating courses presented the experiment in class to their students, who then had two weeks to enroll 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 enrollment 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 treated*. 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 treated*. The remaining 75% of students were given access to streaming only some weeks at random and we label this group the *sometimes treated*. Every week, a varying share of students in this group was given access. In the Spring semester 2017 we randomly assigned weekly access to 50% of the sometimes treated. In the Fall semester 2017 we decided to vary this share between 20% and 80%.¹¹

one or more courses of the first semester with the second cohort.

¹⁰The 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, USA Today and others.

¹¹ 80% in week 3, = 40% in week 4, = 60% in week 5, = 20% in week 6, = 80% in week 7, = 40% in week

At the end of week 2, students were notified by email about the exact sequence of weeks when they were given 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.

3 Theory

In this section we develop a simple model of learning and attendance choice that is useful to guide our empirical analysis. The setup is kept simple and builds on existing literature on the education production function (Hanushek, 1986). We depart 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 (innate) 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.¹³

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.¹⁴ Students value high academic performance as measured by exam grades.¹⁵

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

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

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 the students:

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

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

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

where $\beta_i \in (0, 1)$ represents the innate ability level of student i and $\gamma^j \in (0, \frac{1}{2}]$ is a technological parameter describing how ability affects the returns to effort. This parametrization implies that, given the production technology, high-ability students have higher returns to effort compared to their low-ability peers. It also captures our intuition that 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 size does not allow intense interaction between students and teachers. High-ability students can read the material on a textbook

8, week 9 was mid-term break, = 60% in week 10, = 20% in week 11, = 80% in week 12, = 40% in week 13, = 60% in week 14.

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

¹³We could allow the cost of effort to be larger for less able students, as in some well-known models of education (Spence, 1973), but this would be an unnecessary complication in our setting.

¹⁴In what follows we use the terms *learning technology* and *attendance* or *learning mode* interchangeably.

¹⁵We do not enter the debate on whether exam grades correctly measure learning.

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.¹⁶

Assumptions. We impose the following set of assumptions.

(A1) Classroom attendance is the most efficient learning technology:

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

(A2) The costs of both no attendance and streaming are normalized to zero:¹⁷

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

(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 $[\underline{u}, \bar{u}]$. The shock u is realized before the choice of attendance mode is made.¹⁸

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

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

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:

$$e_i^j = \beta_i^{\gamma^j} \tag{5}$$

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:

$$w_i^j = \frac{\beta_i^{2\gamma^j}}{2} - c^j \tag{6}$$

It is noteworthy that both the optimal level of effort and the associated welfare increase with β_i , the student's ability. However, given the concavity of grades with respect to β_i , low-ability students benefit relatively more from a more efficient learning technology compared to their high-ability peers. In other words, the difference in grades and utility between different modes

¹⁶In smaller classes where interactions are more intense the complementarity between student and teacher abilities would be much more important and make the returns to attendance much higher for the good students.

¹⁷This is a simplifying assumption allowing to limit the number of cases to analyze. Introducing a positive cost of streaming (or 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.

¹⁸The normalization of c^a to zero in the absence of shocks is also a simplifying assumption. Relaxing it would complicate the analysis without changing its main implications.

of attendance is greater for lower ability students. This captures the intuitive idea that students at the upper end of the ability distribution will comprehend the material even without the presence of a professor; 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

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

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.¹⁹

Figure 2 illustrates the functioning of the model and plots indirect utility by ability and attendance modes, with and without shocks. For simplicity, we have set γ^n to 1/2 so that indirect utility when not attending is linear in ability. The figure illustrates that only the high-ability students switch from attending to not attending when exposed to exogenous shocks. Equation 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.

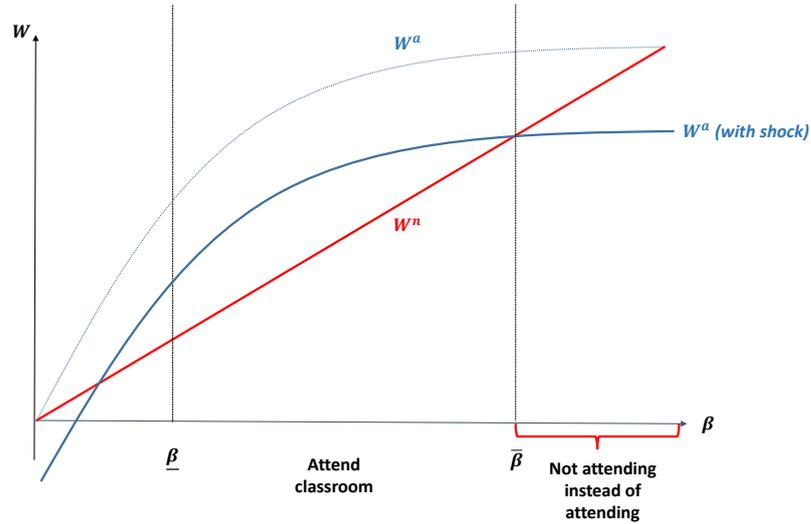


Figure 2: Indirect utility and shocks without streaming

¹⁹ $\bar{\beta}_u$ is defined as $\frac{\bar{\beta}^{2\gamma^a} - \bar{\beta}^{2\gamma^n}}{2} = u$.

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 5, this parametrization is consistent with our data, where almost no student uses the streaming platform every time she has access to it (see Figure 7).

Figure 3 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 constraint $\beta_i > \underline{\beta}$).²⁰ 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, 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.

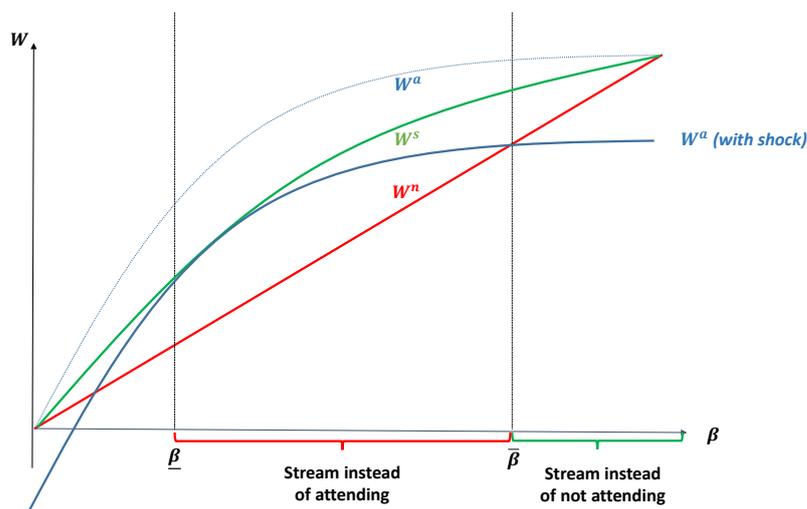


Figure 3: Indirect utility and shocks with streaming

Figure 4 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 document empirically in section 5.

²⁰With smaller shocks the effect on grades would be null for some students at the bottom of the ability distribution.

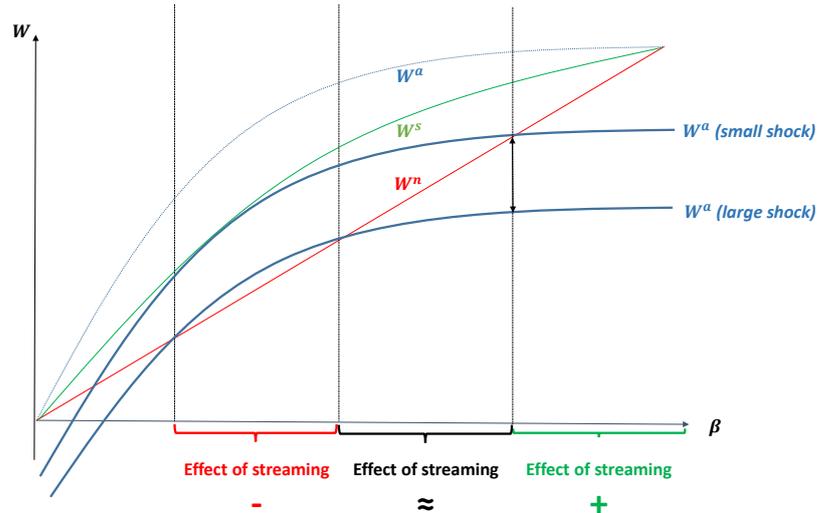


Figure 4: The effect of streaming on grades

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 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.

4 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 individual-level data to assess students' attendance of live streaming based on server logs. Specifically, we used detailed data from the streaming server assessing if, when, and for how long students were logged in.²¹ Of course, we also know which students

²¹In 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, etc. Each of these events are potentially recorded differently depending on the machine (laptop, desktop, tablet or smartphone), the operating system and the browser.

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 number of questions ranging between 16 to 60.²² 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.²³

Next, we have proxy measures of classroom attendance, unfortunately only for the Fall term 2017. We asked a research assistant to take multiple (2 or more) pictures from the back of the auditorium, 15 to 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 the Appendix 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 arithmetic mean taken across all raters.

Finally, 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, as described in Section 3. 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.²⁴ "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.²⁵

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.²⁶ 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

²²Introduction to management has 60 questions, Mathematics 16, introduction to microeconomics 30 (all sections), introduction to macro 30, probability&statistics 16 and human resource management has 30.

²³On average, the material covered in a particular week is useful to answer 2.5 questions.

²⁴We compute the seasonal distributions using measurements for all days from September 2016 to December 2017.

²⁵We have experimented with several variations of these definitions and results change only marginally.

²⁶Source: Office Fédéral de la santé publique, *Sentinella OFSP*.

zero.

4.1 Descriptive statistics

We work with data for 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, with a minimum of 3 and a maximum of 6.

Table 2: Descriptive statistics

	All		Never treated		Randomly treated		Always treated		
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
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	
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 ^a	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	
			Outcomes						
Drop-out ^b	0.295	0.456	0.286	0.453	0.295	0.456	0.308	0.463	
Exam grade ^c	-0.003	0.939	0.047	0.870	-0.026	0.958	0.049	0.923	
Share of correct answers ^d	0.556	0.204	0.568	0.197	0.553	0.204	0.556	0.211	
Observations	1621		252		1119		250		

^a Share above the passing grade.

^b Share of students not taking at least one exam.

^c Share of students streaming at least once.

^c Average normalised grade within each course.

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

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

About 60% of the students are Swiss nationals and the rest are foreigners.²⁷ 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 all of them relative to the system-specific passing grade. On average, our students have a high school leaving grades 33% above the passing grade. Around 35% (43%) have a mother (father) with a college degree. Consistent with the random allocation, differences in all

²⁷The 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.

these descriptive statistics across the three treatment groups are minor and non-significant.²⁸

The lower panel of Table 2 reports descriptive statistics for some outcome variables. We define drop-out 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 out. 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.²⁹ We conducted a series of robustness checks and did not find any effect of the experimental conditions on drop-outs (see Table A-1 in the Appendix).

Exam grades originally range on a scale of 1 to 6 points and we have standardized them to have mean zero and variance one within each course.³⁰ There are some differences across groups, with higher grades for the never-treated and always-treated 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 experimental groups are minor and not significant.³¹

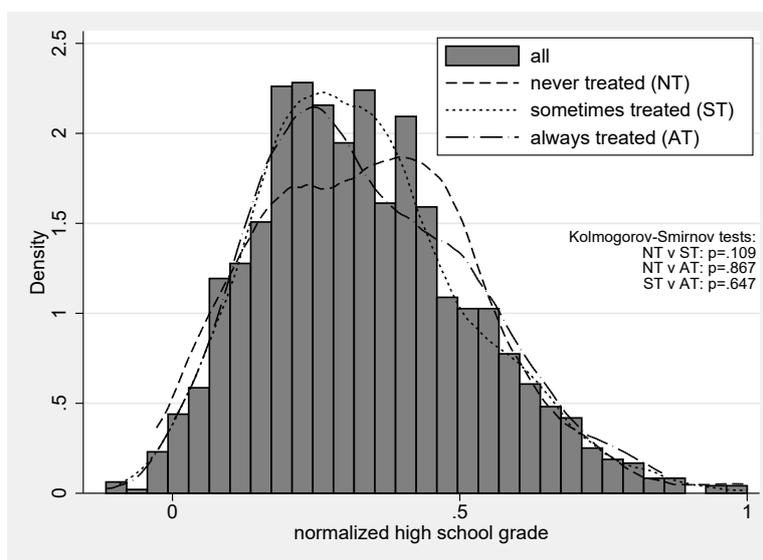


Figure 5: Ability distributions

In Figures 5 and 6 we provide further evidence that the randomization procedure was successful. Figure 5 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 treatment group and we report the results of formal

²⁸We run a full set of mean comparison tests for all variables and all pairs of treatment groups (30 tests in total). Only three such tests (marginally) reject the equality assumption.

²⁹In 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 enroll and drop out during their first year. Some of them often enroll at a *Haute Ecole* or University of Applied Sciences, the vocational tier of higher education.

³⁰There 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.

³¹Most questions have more than two possible answers.

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 experimental groups.

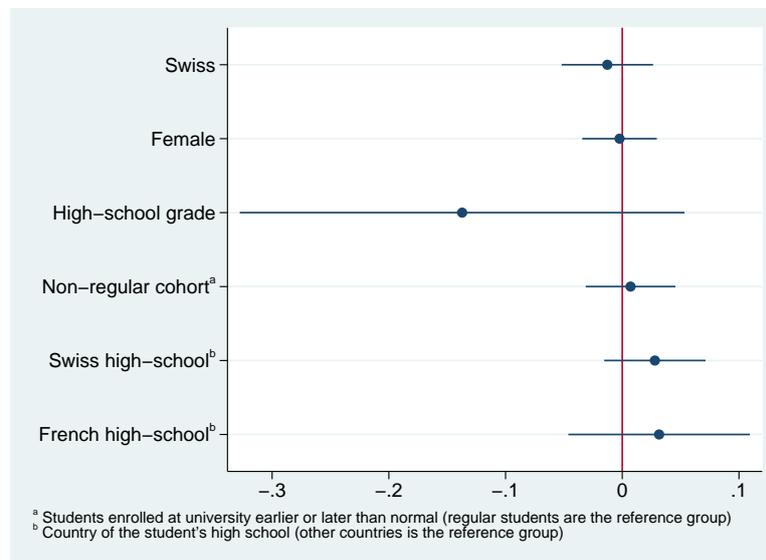


Figure 6: Randomization checks

Randomly allocating students to one of the three treatment groups is only the first step of our randomization procedure. We further randomly assigned the students in the sometimes-treated 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 6 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 of each semester during which lectures were streamed, about 50% of the students had access to the platform.³² 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.³³

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, i.e. the share of students with access who logged into the platform. This share is on average

³²Recall that these averages are taken over all three treatment groups.

³³In the Fall term we varied the share of students with access week by week. See also footnote 11.

Table 3: Weekly assignment and take-up

Week	Spring 2017			Fall 2017		
	Treated ^a	Streamed ^b	Take-up ^c	Treated ^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
9	0.471	0.031	0.067	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
Student-week obs.		12'617			20'944	

^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.

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.

Table 4: Proxy measures of classroom attendance

	Enrolment ^a	Mturk evaluations ^b		
		mean	median	std.dev.
Micro (FR1)	369	226.08	173.89	170.80
Micro (FR2)	513	224.26	186.33	128.04
Micro (EN)	249	134.80	99.14	129.32
Mathematics	486	201.66	165.40	118.57
Management	287	214.40	172.57	128.67

^a Number of students enrolled in the course.

^b Statistics computed over the 252 raters of the pictures of each classroom session.

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.

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

5 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.

5.1 Take-up of the streaming service

As we documented already in Section 4, 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 behavior. 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. This second mechanism is the one we develop in our theory of Section 3 and we show here that it is indeed consistent with the data.

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.

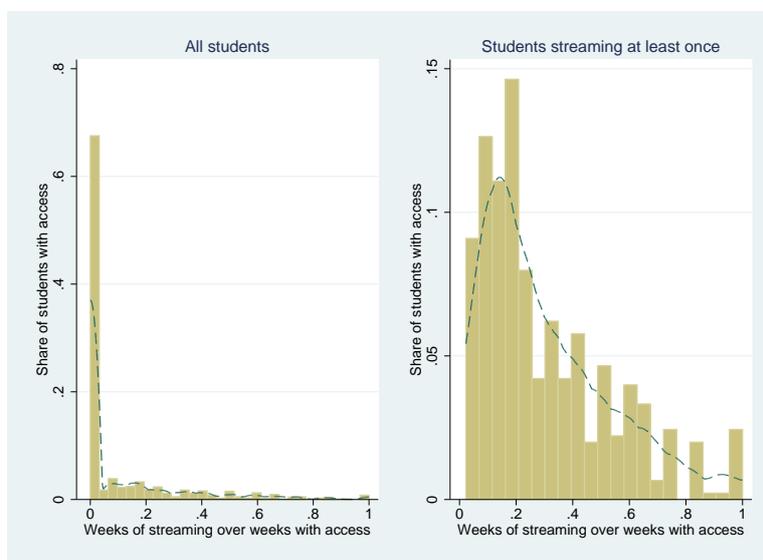


Figure 7: Distribution of cumulative take-up

Our data allow us to compute the distribution of cumulative take-up and we report it in Figure 7, 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. This is

consistent with the idea that students use the service only when random events make the cost of class attendance particularly high.

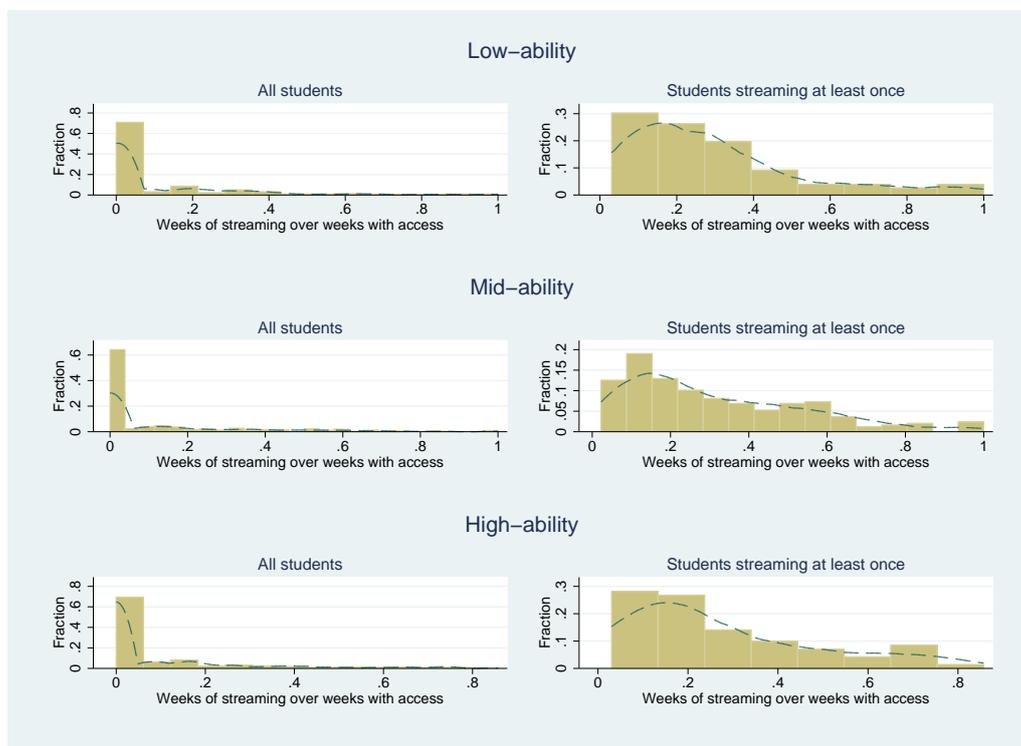


Figure 8: Distribution of cumulative take-up

In Figure 8 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 9 we report the average take-up rate across weeks by percentiles of the distribution of high school grades. We compute take-up as the ratio between the number of students who had access to the streaming platform and the number of those who actually logged in.

Despite some sizable variation between 0 and 20%, there seems to be only 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 enrollment 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-2 in the Appendix.

To further support our theoretical intuition that students use the streaming service only when hit by shocks that increase the cost of class attendance, we have merged our data with information on weather conditions and influenza incidence rates in Geneva (see section 4 for more details). Although weather and health shocks do not exhaust all possible events affecting the cost of attending lectures in person, we believe that they represent a meaningful and rather

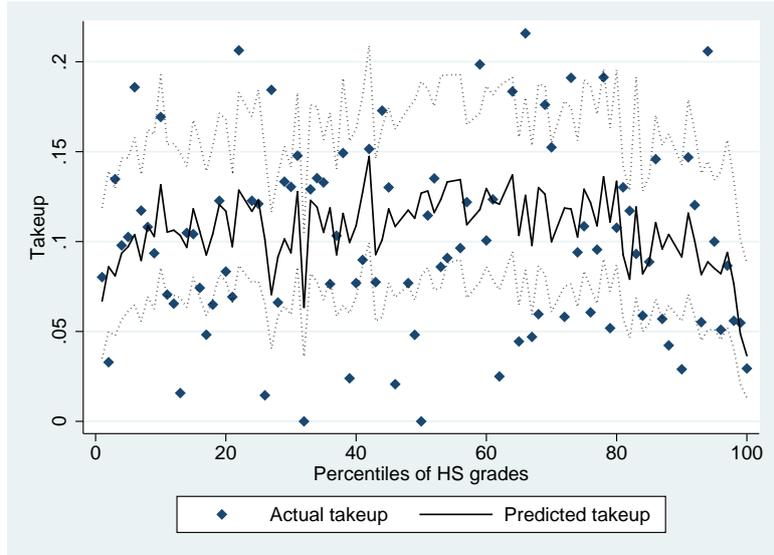


Figure 9: Distribution of weekly take-up by ability

important 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).³⁴

Table 5: 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)

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, High=80th-100th.

For readability, we report in Table 5 predicted probabilities of take-up based on the estimates of this probit model, for different types of students and different weather and influenza

³⁴We only modified the set of controls by replacing the week fixed-effects with season fixed-effects. The weather variables only show too limited variation within weeks to allow separate identification of the week fixed-effects.

conditions. Consistent with our theory, 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. 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 is consistent with the idea that students stream lectures instead of attending in class only punctually. They seem to have a preference for in-class attendance and use live streaming only as a replacement when the cost of going to class is too large. We believe that this is an important and novel result. Other papers in this literature treat the choice of students about the mode of course attendance as an identification nuisance to be controlled for in the attempt to identify a causal effect on performance. Our experimental design allows us to investigate students' choices in details and we show that, under normal conditions, students of all abilities strictly prefer class attendance over streaming. We believe that this is a finding of major importance for understanding the potential impact of distance learning technologies and for the design of related policies.

5.2 Classroom attendance

The theoretical analysis of Section 3 suggests that only some of the students who use the streaming service do so as a replacement of in-class attendance. Hence, we expect the experiment to have only a limited impact on attendance.

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 4 for details). Unfortunately, this information is far from ideal. We only have 42 observations, one for each lecture participating in the experiment in the Fall 2017, and the Mturk evaluations are noisy, as documented in Table 4.

Eventually, we estimate regression models of the following type:

$$attendance_{cw} = a_1^1[\# treated]_{cw} + a_c^1 + e_{cw}^1 \quad (8)$$

$$attendance_{cw} = a_1^2[\# streamed]_{cw} + a_c^2 + e_{cw}^2 \quad (9)$$

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 ; $[\# treated]_{cw}$ and $[\# streamed]_{cw}$ 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 intention-to-treat effect (ITT), and a_1^2 , which measures the average treatment effect on the treated (ATT). Given the potential endogeneity of students' take-up decisions, we estimate equation 9 using $[\# treated]_{cw}$ as an instrument for $[\# streamed]_{cw}$ and by doing so we clearly identify a local effect on the unobservable population of compliers. Notice, however, that under the model interpretation that students use streaming only when hit by random shocks (an interpretation that is consistent with the empirical evidence in the previous section 5.1), all students are 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 8 and 9 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 6 and, despite the small sample size and the noisy

Table 6: Treatment effects on classroom attendance

	OLS	OLS	IV
Number of treated students	-0.079** (0.037)	-	-
Number of students streaming	-	-0.431 (0.260)	-0.654** (0.313)
Obs.	42	42	42
Mean	153.6	153.6	153.6
Sd	37.1	37.1	37.1

The dependent variable is the median evaluation of the classroom pictures by the MTurk evaluators. All observations are weighted by the inverse of the standard deviation of the individual evaluations.

variables, 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.

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 10 students logging into the platform about 6 of them do not show up in class). Following our theoretical framework, this result can be rationalized recalling that for high-ability students the counterfactual to streaming is not attending the classes. This mechanism also explains why the estimated effect of streaming on attendance is lower than average take up (10%).

Simple back-of-the-envelope calculations also seem to be consistent with this interpretation. Consider students in the top 20% of the ability distribution (which is our definition of high-ability) as those who indeed substitute no attendance with streaming when they have access to it. All other students substitute in-class attendance with streaming. Then, if the take-up rate is 10% and 80% of the students use streaming instead of going to class, then we should exactly find an ITT of approximately 0.08.³⁵

Interestingly, the IV estimates are larger than OLS (although not significantly) suggesting that students might stream more when the class is more crowded.

5.3 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 a dummy indicator for answering correctly 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 4 for details). We use this dependent variable y_{icw} in linear probability models with the following specifications:

$$y_{icw} = \alpha_1^1[\textit{treated}]_{iw} + \alpha_2^1 X_i + \alpha_c^1 + \theta_w^1 + \eta_i^1 + \epsilon_{icw}^1 \quad (10)$$

$$y_{icw} = \alpha_1^2[\textit{streamed}]_{icw} + \alpha_2^2 X_i + \alpha_c^2 + \theta_w^2 + \eta_i^2 + \epsilon_{icw}^2 \quad (11)$$

³⁵According to the theory, take-up is equal to probability of the shock occurring (p) times the probability that the shock is large enough to induce a change in the choice of attendance mode.

where $[treated]_{iw}$ is a dummy indicator equal to one if student i was assigned to have access to the streaming platform in week w . Similarly, $[streamed]_{icw}$ 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 enrollment cohorts (early, late and regular enrollers) and three nationalities of the high school diploma (Swiss, French and other nationalities).³⁶ $\{\alpha_c^1, \alpha_c^2\}$ 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. Obviously, when we estimate equations 10 and 11 with student fixed effects we cannot identify the parameters $\{\alpha_2^1, \alpha_2^2\}$. Finally, $\{\epsilon_{icw}^1, \epsilon_{icw}^2\}$ are residuals.

As with equation 9, we estimate also equation 11 using $[treated]_{icw}$ as an instrument for $[streamed]_{icw}$. 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).³⁷

Table 7: Treatment effects on exam performance

	Random effects			Fixed effects		
	ITT	ATT	OLS	ITT	ATT	OLS
All students	0.001 (0.005)	0.006 (0.042)	-0.006 (0.011)	0.003 (0.005)	0.023 (0.043)	-0.004 (0.010)
<i>By ability group^a</i>						
Low	-0.019** (0.009)	-0.181* (0.098)	-0.019 (0.027)	-0.019** (0.009)	-0.178* (0.102)	-0.024 (0.027)
Mid	0.000 (0.006)	0.001 (0.047)	-0.002 (0.013)	0.000 (0.006)	0.001 (0.050)	-0.005 (0.013)
High	0.025** (0.010)	0.249** (0.110)	-0.008 (0.026)	0.023** (0.011)	0.241** (0.121)	-0.010 (0.028)
Obs. ^b	23'766	23'766	23'766	23'766	23'766	23'766
<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

^a Low/Mid/High based on percentiles of the distribution of high school grades. Low=1st-20th, Mid=20th-80th, High=80th-100th.

^b One observation per student-course-week.

Results of the estimation of equations 10 and 11 are reported in Table 7. 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.³⁸

Results indicate that on average there is no detectable effect of the experimental intervention, nor of actual usage of the streaming platform. Consistent with the theoretical analysis of Section 3, once we look at effects by ability groups we uncover large heterogeneity, with a

³⁶The ability groups are defined as in Figure 9, 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.

³⁷Full results of the first stage regressions, including those estimated on the three ability sub-groups of students, are reported in Table A-3 in the Appendix.

³⁸These heterogeneous effects are obtained by interacting the variables $[treated]_{icw}$ and $[streamed]_{icw}$ (both in the main regressions and in the first-stage of the IV model) with dummies for the three ability groups.

sizable negative intention-to-treat effect on the low-ability students and a positive effect on the high-ability students. 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.

The magnitudes of these estimates are sizable. 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 probability of answering exam questions correctly 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 -20 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 5.1. We believe that these findings are consistent with our interpretation that students use the platform only when the cost of in-class attendance is too high due to random shocks. Hence, if the streaming service was scaled up and offered to all students for the entire term, we should expect an increase in the probability of correct answer 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 random.

Following the theoretical results developed in section 3, we interpret the heterogeneity of effects across ability groups as being generated by heterogeneity in counterfactuals. High-ability students substitute streaming to not attending, with a positive impact on their grades. Low-ability students substitute streaming to attending the class and suffer a decrease in grades.

Notice also that both the ITT and the ATT are precisely estimated around zero for students with high school grades in the middle of the distribution, which is also consistent with our theory of Section 3.

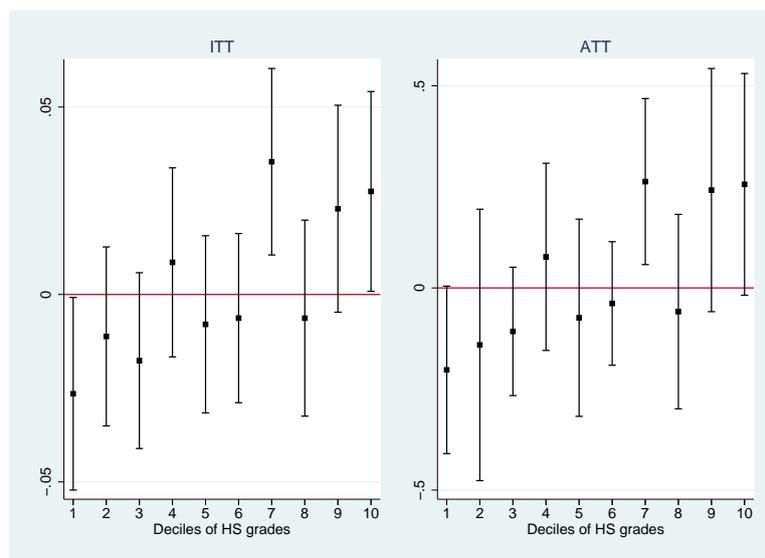


Figure 10: ITT and ATT by student ability

Figure 10 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 10 and 11 where the variables $[treated]_{icw}$ and $[streamed]_{icw}$ are interacted with a full set of dummies for the student’s decile in the distribution of the high school leaving grades.³⁹ 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 7 is not just a mere statistical artifact.

It is worth noticing that Figure 10 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 learning effect of distance learning technologies increases with the student’s ability. It 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 US. Hence, they did not observe students in the upper part of the ability distribution for whom the effect of online learning turns positive. In 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.

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;
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 create a theoretical framework that successfully rationalizes all these findings and is consistent with other pieces of empirical evidence. The model predicts 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.

We believe that our results represent the most comprehensive set of findings to date on how distance learning technologies affect attendance and exam performance in higher education. We further provide insights on the heterogeneity of the effects across students of different abilities. We exploit experimental variation with a large sample, we study take-up and attendance decisions and we provide a theory that is consistent with the findings.

³⁹The estimates (and confidence intervals) in Figure 10 are produced with the specification using student random effects. Using fixed effects does not change the results in any meaningful sense.

These results also provide essential information 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 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 reduce class size.

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.

Our analysis is specific to a very common learning environment, namely large introductory classes with many students and rather standardized content. Hence, despite being a single case study it easily generalizes to many similar settings. However, it is also fair to say that it is unlikely to generalize 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, thus reversing the implications of our model. 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.

References

- William T. Alpert, Kenneth A. Couch, and Oskar R. Harmon. A randomized assessment of online learning. *American Economic Review: Papers & Proceedings*, 106(5):378–382, 2016.
- Abhijit V. Banerjee and Esther Duflo. (dis)organization and success in an economics mooc. *American Economic Review*, 104(5):514–18, May 2014.
- Eric P. Bettinger, Lindsay Fox, Susanna Loeb, and Eric S. Taylor. Virtual classrooms: How online college courses affect student success. *American Economic Review*, 107(9):2855–2875, 2017.
- William G. Bowen, Matthew M. Chingos, Kelly A. Lack, and Thomas I. Nygren. Interactive learning online at public universities: Evidence from a six-campus randomized trial. *Journal of Policy Analysis and Management*, 33(1):94–111, 2014.
- Byron W. Brown and Carl E. Liedholm. Can web courses replace the classroom in principles of microeconomics? *American Economic Review*, 92(2):444–448, May 2002.
- Dennis Coates, Brad R. Humphreys, John Kane, and Michelle A. Vachris. “no significant distance” between face-to-face and online instruction: evidence from principles of economics. *Economics of Education Review*, 23:533–546, 2004.
- David J. Deming, Claudia Goldin, Lawrence F. Katz, and Noam Yuchtman. Can online learning bend the higher education cost curve? *American Economic Review*, 105(5):496–501, May 2015.
- David J. Deming, Noam Yuchtman, Amira Abulafi, Claudia Goldin, and Lawrence F. Katz. The value of postsecondary credentials in the labor market: An experimental study. *American Economic Review*, 106(3):778–806, March 2016.
- Maya Escueta, Vincent Quan, Andre Joshua Nickow, and Philip Oreopoulos. Education technology: An evidence-based review. *NBER Working Paper Series*, (23744), 2017.
- David Figlio, Mark Rush, and Lu Yin. Is it live or is it internet? experimental estimates of the effects of online instruction on student learning. *Journal of Labor Economics*, 31(4), 2013.
- Eric A. Hanushek. The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24(3):1141–1177, 1986.
- Ted Joyce, Sean Crockett, David A. Jaeger, Onur Altindag, and Stephen D. O’Connell. Does classroom time matter? *Economics of Education Review*, 46:64 – 77, 2015.
- Cory Koedel, Rajeev Darolia, Paco Martorell, Katie Wilson, and Francisco Perez-Arce. Do Employers Prefer Workers Who Attend For-Profit Colleges? Evidence from a Field Experiment. Working Papers 1411, Department of Economics, University of Missouri, June 2014.
- Michael S. McPherson and Lawrence S. Bacow. Online higher education: Beyond the hype cycle. *Journal of Economic Perspectives*, 29(4):135–154, 2015.
- Michael Spence. Job market signalling. *Quarterly Journal of Economics*, 87(3):355–374, 1973.

U.S. Department of Education. Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies. Technical report, Washington, D.C., 2010.

Ludger Woessmann. The Importance of School Systems: Evidence from International Differences in Student Achievement. *Journal of Economic Perspectives*, 30(3):3–32, Summer 2016.

Di Xu and Shanna Smith Jaggars. 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(3):360–377, 2011.

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

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

Table A-1: Effect of the experiment on dropout

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

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 column 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 panel 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 Low-Mid-High based on percentiles of the distribution of high school grades. Low=1st-20th, Mid=20th-80th, High=80th-100th.

^b Treatment variable is the number of weeks with access during the term.

^c Takeup= 1 if student streamed the course at least once during term.

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

Table A-2: 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

^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-3: 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

^a Low-Mid-High based on percentiles of the distribution of high school grades. Low=1st-20th, Mid=20th-80th, 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 treatment 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.

Appendix B: IV estimates based on commuting distance

For comparison with the existing literature, in this subsection we implement with our data the common instrumental variable strategy based on commuting distance to college that is used in Bettinger et al. (2017); Coates et al. (2004); Xu and Jaggars (2013), three influential papers asking research questions very closely related to ours.

We follow in particular the approach of Bettinger et al. (2017). They use data where each student is observed in multiple exams and propose as instrumental variable the interaction of a binary indicator for courses offered online or in-person and distance from home to campus.

In order to be able to replicate their identification strategy, we have augmented our data with the grades obtained by the students in our main sample in compulsory courses that were not part of our experiment. There are 13 such courses, which students take in their first, second or third term. We can then replicate the approach of (Bettinger et al., 2017) by defining the courses of our experiment as those that were offered online, at least in part, and all others as those offered in person.⁴⁰

Contrary to our main analysis, which exploits variation within-weeks, this IV strategy can only be implemented with observations at the student-course level. In addition, it requires a measure of distance from home to campus. The administrative archives contain information about the addresses of the students at two points in time: at enrolment and in the year 2017, when the experiment was implemented and the data were extracted. Normally, the address at enrolment is the home address where students live or used to live with their parents, whereas the current one is where students reside during their studies.⁴¹

Table B-1: Distance from home to campus by ability group

Ability group ^a	Prop. movers	Distance at enrolment ^b	Current distance ^b	Distance change
Low	0.095	29.474	26.964	-2.511
Mid	0.169	53.341	35.209	-18.132
High	0.282	68.165	39.003	-29.162
Total	0.178	51.752	34.389	-17.363

^a Low/Mid/High based on percentiles of the distribution of high school grades. Low=1st-20th, Mid=20th-80th, High=80th-100th.

^b Distance from home to campus in kilometers (straight line).

We fed all addresses to Google Maps and computed the direct straight line in kilometres from home to campus.⁴² To reduce the influence of abnormal values, we censored the distance measure at 200 kilometres.⁴³

Unfortunately and contrary to the setting of (Bettinger et al., 2017), distance from college appears to correlate with important students' observables, namely ability. This is evident from Table B-1. Overall, about 18% of students changed residence since enrolment but the proportion clearly increases with ability, from around 10% among the least able students to almost 30% among the most able ones. In columns 2 and 3 we observe the same pattern for commuting

⁴⁰We additionally code the experimental courses as not offered online for students who were assigned to the never-treated group, as for these students the courses were effectively not available in online mode.

⁴¹For some students no valid address is available and we loose 220 observations, leading to a sample of 1'401 students that can be used for this analysis.

⁴²We have also computed commuting times and results do not change substantially when using this alternative measure of distance as instrumental variable.

⁴³At enrolment, about 20% of the sample has distance above 200 km. For current addresses, the proportion is 9.8%.

distance at enrolment and at the time of the experiment. The last column of Table B-1 looks at the actual change in commuting distance for the movers. On average students move closer to the university but the most able are the ones who move the most.⁴⁴

Bettinger et al. (2017) provide evidence that in their context distance to college does not violate the required exogeneity assumption, whereas the descriptive evidence in Table B-1 suggests that this assumption might actually fail in our setting. Despite of this concern, it is still interesting to compare our findings with comparable results produced using this instrumental variable strategy. In Table B-2 we report estimates of α_1^{iv} from the following model:

$$y_{ic} = \alpha_1^{iv} stream_{ic} + \alpha_2^{iv} X_i + \alpha_3^{iv} dist_i + \alpha_c^{iv} + \epsilon_{ic}^{iv} \quad (12)$$

where y_{ic} is the final grade of student i in course c (standardised within each course); $stream_{ic}$ is the share of weeks in the term when student i used the streaming platform for course c ; $dist_i$ is a measure of commuting distance for student i . X_i is our usual set of individual controls; α_c^{iv} is a course fixed effect and ϵ_{ic}^{iv} is the residual (including a student random effect).

Table B-2: Instrumental variable estimates of final grades

Instrument ^b =	Dist. at enrolment ^c	Current distance ^c	Randomisation ^d
all students	-1.236 (4.477)	0.514 (2.220)	0.022 (0.481)
F-stat 1st stage	3.3	20.5	454.1
<i>By ability group^a</i>			
Low	-2.741 (3.264)	-1.799 (2.726)	-0.638 (0.936)
Mid	-2.559 (4.130)	0.429 (2.297)	-0.359 (0.511)
High	1.993 (4.827)	5.060 (4.371)	1.598** (0.804)
Obs.	5153	5153	5153
<i>Descriptive stat. of the dep. variable</i>			
Mean	0.015	0.015	0.015
Sd	0.972	0.972	0.972

^a Low/Mid/High based on percentiles of the distribution of high school grades. Low=1st-20th, Mid=20th-80th, High=80th-100th.

^b The endogenous variable is the streaming frequency (the number of weeks with at least one connection to the streaming platform of the course over the number of weeks in the term).

^c Instrument is the interaction between distance (from home to campus in kilometers) and a dummy indicator if the course was offered via live streaming to the student (i.e. it participated in the experiment and the student had access to streaming).

^d Instrument is the treatment frequency (the number of weeks with access over the number of weeks in the term).

In the first two columns of Table B-2 we report estimates of α_1^{iv} obtained instrumenting $stream_{ic}$ with the interaction between $dist_i$ and a dummy indicator for courses that participated in our experiment (which is subsumed in the course fixed effects when not interacted). In column one we use commuting distance measured at enrolment and in column two the one measured during the experiment. Following the same structure as Table 7, the upper panel of

⁴⁴The correlations of the variables in Table B-1 with ability resist also when conditioning on our usual set of student characteristics.

the table shows estimates for the entire sample whereas the lower panel shows heterogeneous effects for the three ability subgroups.

Results are broadly consistent with the evidence we presented in Section 5.3, although the smaller sample sizes do not always allow reaching conventional levels of statistical significance. Attending lectures via streaming has positive effects on high-ability students, declining along the distribution towards zero to negative effects for the least able ones. Using current distance instead of distance at enrolment improves the relevance of the instrument and also shifts upwards the distribution of the effects.

For comparison, in the third column of Table B-2 we report estimates of equation 12 (with and without interactions with the ability groups) using an alternative instrument based on the experimental variation in our data. Specifically, we define the instrument as the share of weeks in the term when the student was randomly assigned to having access to the streaming platform. The results are consistent both with the previous columns and with our main findings in Section 5.3.

Overall, the results of this exercise suggest that our findings are consistent with those in previous papers and, importantly, that they are not specific to a particular identification strategy.