Stanford cepa | Center for Education Policy Analysis

CEPA Working Paper No. 18-11

The Effects of Blended Online Learning in Higher Education STEM Courses: Experimental Evidence from Mongolia

AUTHORS

Jamie Johnston
Stanford University

ABSTRACT

In low- and middle-income countries, institutions of higher education are turning to online models of instruction to reduce costs and broaden their educational reach. While a growing body of causal research can speak to the effectiveness of online models in the United States, there is little rigorous evidence about the use of online models in lower income countries. To fill this gap in the research, I use a randomized design to examine the effectiveness of a blended model in undergraduate STEM courses in Mongolia. On average, students assigned to the online instructional format withdraw from courses at a higher rate; this finding is not observed among the highest achieving students, suggesting lower-ability students may encounter barriers to persistence under new online learning models. Nevertheless, overall course performance is equivalent between treatment and control, suggesting the online model may be as effective as face-to-face instruction at a lower cost.

VERSION

May 2018

Suggested citation: Johnston, J. (2018). The Effects of Blended Online Learning in Higher Education STEM Courses: Experimental Evidence from Mongolia (CEPA Working Paper No.18-11). Retrieved from Stanford Center for Education Policy Analysis: http://cepa.stanford.edu/wp18-11

The Effects of Blended Online Learning in Higher Education STEM Courses: Experimental Evidence from Mongolia

Jamie Johnston[1,2]

Working Paper

May 2018

ABSTRACT: In low- and middle-income countries, institutions of higher education are turning to online models of instruction to reduce costs and broaden their educational reach. While a growing body of causal research can speak to the effectiveness of online models in the United States, there is little rigorous evidence about the use of online models in lower income countries. To fill this gap in the research, I use a randomized design to examine the effectiveness of a blended model in undergraduate STEM courses in Mongolia. On average, students assigned to the online instructional format withdraw from courses at a higher rate; this finding is not observed among the highest achieving students, suggesting lower-ability students may encounter barriers to persistence under new online learning models. Nevertheless, overall course performance is equivalent between treatment and control, suggesting the online model may be as effective as face-to-face instruction at a lower cost.

1 Introduction

Around the globe, institutions of higher education are taking their classrooms online to reduce costs and broaden access. Supporting this transition, development agencies and large philanthropic donors are channeling funds into education technology and online learning interventions in low- and middle-income countries (Cheney, 2017). While much of the enthusiasm around online learning in developing countries still centers around massive open online courses (MOOCs), funders are also increasingly directing attention toward smaller and more personalized online learning (Cheney, 2017; Robertson, 2015).

Even so, a growing body of causal research suggests that online substitutes for traditional in-person instruction yield inferior student outcomes (Bettinger et al., forthcoming; Alpert et al., 2016; Hart, et al., 2016; Streich, 2014; Figlio et al., 2013; Xu and Jaggers,

1 Stanford University, Postdoctoral Fellow, jamie.johnston@stanford.edu
2 I gratefully acknowledge the financial assistance of the Freeman Spogli Institute for International Studies (FSI) Global Development and Poverty Initiative, the Institute of Education Sciences US Department of Education (grant R305B140009), and the Stanford University Vice Provost for Online Learning.

2013). However, students have been shown to learn better through blended models of instruction that combine online interactions with face-to-face instruction than they do through purely remote instruction (Alpert et al., 2016; Bowen et al., 2014), and face-to-face time appears to be an important factor in student learning (Joyce et al., 2015).[3]

To date, however, experimental studies of online learning have largely investigated applications at four-year universities in the United States (Escueta et al., 2017). As in the U.S., online instruction is spreading in low- and middle-income countries (Cheney, 2017), but no experimental studies measure the effectiveness of online learning in those less resourced countries[4]. To fill this gap in the research, this study employs a randomized design to estimate the effectiveness of a blended online model piloted in undergraduate STEM courses in a lower-middle income country, Mongolia.

Educational institutions have increasingly moved instruction online in an effort to reduce costs and increase accessibility, and new evidence suggests they may be justified in doing so. Deming et al. (2015) observe that through reduced labor costs and economies of scale, institutions leveraging online delivery of instruction may be able to lower costs and likewise tuition, holding demand-side implications for access. On the supply side, increasing online class sizes comes with little increase in operational cost. Moreover, with respect to student outcomes, online settings may be less sensitive to class size increases compared to in-person settings (Bettinger et al., 2017). Indeed, new empirical evidence demonstrates that

---

[3] Online instruction is often categorized in two forms. In solely online settings, interactions between students and instructor(s) always take place remotely and through virtual means, generally through internet connection. In blended settings, students and instructor(s) spend at least some amount of time in a face-to-face setting, and instruction is supplemented by online instructional videos or other digital learning tools.

[4] A National Bureau of Economic Research (NBER) review of education technology interventions (Escueta et al., 2017) names 7 RCTs that compare online versus face-to-face: Alpert et al., 2016; Bowen et al., 2014; Figlio et al., 2013; Heppen et al., 2012; Joyce et al., 2015; Keefe, 2003; Poirier and Freeman, 2004. Zhang, 2005, all of which were conducted in the United States. Additionally, I searched the AEA list of registered RCTs, World Bank publications, and the NBER working paper series and find no RCT or other quasi-experimental studies examining the effectiveness of online learning models compared to traditional instruction in a low- or middle-income country.

online programming can dramatically increase the number of students trained (Goodman et al., 2016).

In lower-income countries, where instructors' pedagogical expertise and knowledge of technical content may be limited, the promise of online instruction is especially attractive. On average, low- and middle-income countries have lower levels of human capital (Barro & Lee 1993, 1996, 2001). Selective outmigration of experts might also lead to a smaller subset of faculty in institutions of higher education. Through online content developed within countries, institutions could widen the reach of the available experts. Moreover, as argued through a stylized model by Acemoglu et al. (2014), lower-skilled teachers can leverage the comparative advantage of more skilled teachers (within and outside the country) through online resources to distribute educational resources more equally within and across societies.[5]

Importantly, the potential for web-based resources to improve national education systems hinges on whether online models of instruction are effective in producing student learning. Although a number of non-causal studies tout the success of online models (Means et al., 2010), a small but growing body of rigorous causal research suggests that simply using online instruction to replace traditional face-to-face instruction results in inferior student outcomes. Using quasi-experimental designs, Bettinger et al. (forthcoming), Hart et al., (2014), Streich (2014), and Xu and Jaggers (2011, 2013) find that students taking courses online score lower on assessments and are less persistent in their courses as compared to students taking courses in a traditional face-to-face format. The two published randomized controlled trials that examine the effectiveness of purely online instruction (i.e., no face-to-face component) find that students in the online settings perform worse than do students

[5] Indeed, methods of distance instruction have long been used (e.g. mail correspondence, radio, and video recordings) in developing countries to expand curricular access; however, there is, to my knowledge, no causal evidence of its effectiveness compared to in-person instruction that covers the same content.

taking the same courses in traditional face-to-face settings (Alpert et al., 2016; Figlio et al., 2013).

Students appear to learn better from blended models that combine online instruction with in-person support. The experimental studies that compare the outcomes of blended and traditional formats (Alpert et al., 2016; Bowen et al., 2014; Lovett et al., 2008) find that students perform equivalently in both settings.[6] Alpert et al. (2016) also find that students in a blended setting outperform students in an online-only setting. These studies suggest that face-to-face interaction is an important aid in student learning. Joyce et al., (2015) confirm this hypothesis by using a randomized design to examine the impact of increased face time; they find that students in a blended setting with two hours of in-person instructor interaction significantly outperform those with only one hour of instructor face time.

Although these studies are useful first steps for understanding the effectiveness of online learning, the extant studies are limited in scope and are not necessarily generalizable to low-resourced settings in developing countries where students face different challenges. The aforementioned experiments comparing online learning to traditional learning were all conducted with undergraduate volunteers at four-year universities in the United States (Alpert et al., 2016; Bowen et al., 2014; Figlio et al., 2013). Furthermore, the studies examine online versions of just two introductory-level courses: microeconomics (Alpert et. al, 2016; Figlio et al., 2013) and statistics (Bowen et al., 2014). Two of the studies (Figlio et al., 2013; Bowen et al., 2014) had participation rates of under 25 percent (measured as a percentage of students recruited to participate in the study), and therefore their results might not be generalizable to

[6] Using a smaller sample of students (N=68), Lovett, Meyer, and Thille (2008) also provide early evidence on the effectiveness of the same hybrid online statistics course evaluated by Bowen et al. (2014). The experimental evidence from Bowen et al. on a larger sample (N=605) confirm Lovett et al.'s finding that the hybrid model is equally effective as a traditional face-to-face model.

the full corpus of course registrants.[7]

The need for evidence on the effectiveness of online learning in lower-income countries is growing – not only are citizens of low- and middle-income countries accessing educational resources at increasingly higher rates, they may face barriers not encountered by learners in wealthier countries. Citizens of low- and middle-income countries now comprise the majority of MOOC users worldwide (Garrido et al., 2016); yet they score substantially lower and are less likely to persist in their courses compared to counterparts in wealthy countries (Kizilcec and Halawa, 2015). Obstacles such as access, language, and computer literacy, as well as barriers related to social identity threat (i.e., lower self-efficacy caused by identity-related anxieties) may limit their potential for learning (Liyanagunawardena et al., 2013; Kizilcec et al., 2017).

Blended models might be especially useful in lower-income countries where face-to-face support could mitigate some of these challenges. Although blended models of instruction have not been studied directly in these nations, a number of studies show that computer-assisted learning (CAL) interventions can improve learning outcomes among students in low-income nations (Banerjee et al., 2007; He et al., 2007; Lai et al., 2015; Muralidharan et al, 2016). Furthermore, a systematic review suggests that CAL interventions are more effective in developing countries than they are in developed ones (Bulman & Fairlie, 2016). While the literature on CAL interventions suggest that online supplements to higher education might be especially effective in developing countries, they leave many questions unanswered. Most of these studies were conducted in primary and secondary schools, not universities; moreover, the online content was largely consumed at school during class time or during after-school programming. These studies therefore do not necessarily capture the effectiveness of online

[7] The participation rate of Alpert et al. (2016) was not reported.

lessons that students typically consume independently and off campus, as they generally would in university contexts.

To fill these gaps in the literature, I employ a randomized design to estimate the effectiveness of a blended model of online learning implemented at a public university in Mongolia. The university piloted the model in seven STEM (Science Technology Engineering and Math) courses that comprise the core curriculum for undergraduate engineering students. In conjunction with a university production team, each instructor developed online videos that presented the material covered in lecture during each of the 16 weeks of the semester. Faculty taught two concurrent sections: 1) a control section taught solely through face-to-face instruction, as the course had been taught in previous years (control); and 2) a blended treatment section in which students received access to online videos and also met with instructors in person for roughly half the time control students met with instructors. Specifically, I address the primary research question – what is the effect of assignment to the blended model on the following academic outcomes: persistence in the course, course grade, persistence in program, and course grades in the two years post treatment?

I find that students assigned to the treatment condition had a higher course withdrawal rate; this higher course withdrawal rate was driven by the lowest-achieving students, suggesting an initial resistance to the new format among the most vulnerable students. However, overall I find no impact on students' overall course score, suggesting learning was comparable among treatment and control groups. This result is robust to bounding to account for the differential withdrawal among the treatment group. In the long-run, I find no difference in course completion. Transcript data collected two years after the intervention reveal that the treatment and control groups displayed equivalent passing rates in the experimental courses.

This paper is organized as follows. Section 2 provides background on the study setting and experimental design. Section 3 describes the data collection and estimation strategy. Section 4 describes the main results. Section 5 unpacks compliance and course experience. Section 6 concludes.

2 Study Setting and Experimental Design

2.1 Study Setting

Mongolia is a compelling context in which to study online learning for two reasons. First, government reforms aimed at increasing primary and secondary school enrollment have shifted financing away from tertiary education at the same time demand for higher education has surged (UNESCO, 2012). While public universities have therefore increasingly relied on student fees to cover their operating costs, they recognize that their student bodies cannot afford steep increases in tuition. Hence, Mongolian institutions of higher education have been seeking out more cost-effective modes of instruction to meet increased student demand without significantly raising tuition for some time (Sodnomtseren, 2002).[8]

Second, Mongolia's sparse population density makes it an interesting case on how online models could improve the quality of instruction in a country with a large rural population. The majority of Mongolia's tertiary institutions (and all of its selective institutions) are located in Ulaanbaatar. Students living outside the capital must relocate if they wish to pursue high quality higher education – a phenomenon not uncommon in lower income countries where elite institutions are generally located in larger cities and capital cities (Altbach, 2009).

[8] Conversations with university administrators involved with the study confirm this continues to be the case.

The current study took place at a large public university in Ulaanbaatar, the capital of Mongolia. The participating university is one of several selective Mongolian institutions of higher education, and it draws students from across the country. If an online model proved feasible in settings like Mongolian satellite campuses, it could substantially improve the quality of instruction for rural students who currently lack access to well-trained instructors. This study can shed light on the potential feasibility of implementing blended learning models for higher education in similar settings. In subsequent years, the participating university plans to make online options available in satellite campuses outside Ulaanbaatar.

2.2 Blended Learning Pilot

Engineering faculty at the university identified seven courses in which they wished to pilot a blended online model (see Table 1). The faculty members identified these courses as ideal candidates because they are required by multiple majors and therefore are in high demand among students. Several of these courses are taught multiple times per year (i.e., in both the fall and spring semesters). Because the basic content of the courses remains constant from year to year and because the courses are taught each year by the same faculty members, transforming the courses' lectures into online content would reduce the amount of time faculty spend re-teaching lectures every year.

The faculty implemented a "flipped-classroom" approach through which lecture content would be delivered online and face-to-face time with students would be reduced and restructured to a question-and-answer style discussion section during which they could offer more personalized support to students. Under the traditional model, faculty deliver one 90-minute face-to-face lecture each week. Under the flipped model, faculty met face-to-face with students for one 50-minute discussion section each week. The reduced face-to-face time was also intended to allow faculty to reallocate their time toward research, mentorship of graduate students, and administrative duties in the university.

[Table 1]

The faculty members who taught the selected courses created video content intended to replicate the lectures they delivered in person over the 16-week semester. The videos posted online consisted primarily of a recording of the professor lecturing with power-point slides in the background. Some videos also contained laboratory demonstrations similar to those performed during lectures. Online videos were made available to students via an open-source learning management system managed by the university.[9]

The pilot was conducted over two semesters, the Spring and Fall semesters in 2015. All seven courses were offered during the first semester and three were offered during the second semester. The two courses with the largest enrollment (Electronic Fundamentals and Engineering Mathematics) are regularly taught in two sections by two professors each semester. To ensure consistency across instruction, for each course, these professors offered the same assignments and exams and collaborated on creating video content.[10] In total, the study ran for two semesters and included 7 courses taught by 10 professors; the study followed 14 unique course-semester-professor combinations (details in Table 1).

Each professor taught two concurrent sections: 1) a traditional face-to-face lecture section and 2) a blended online section. Students enrolled in traditional classes—the face-to-face lecture section—attended one 90-minute lecture each week as in previous years. Students enrolled in the blended classes received access to weekly video lectures and attended one 50-minute session with the professor or a teaching assistant each week. The

[9] Appendix Figure 1 shows a screenshot of the platform relaying a pre-recorded lecture online.
[10] I use section fixed effects despite course content being similar to address potential differences across sections.

shortened lecture session for blended class participants were not intended to reiterate the material covered in the online lectures, but rather to serve as a discussion section in which professors could provide guidance and supplementary support through discussion, problem solving, and assignment feedback. Only students assigned to the blended format were given login credentials to access the lectures posted online. They could access online content using the university's computer lab, their personal computers, and/or their smart devices.[11]

2.3 Experimental Design

Field surveyors recruited students to participate in the blended classes during the first two weeks of the semester by offering a modest financial incentive (the reduction of a half credit's worth of tuition, equivalent to roughly $11 USD). Field surveyors provided prospective students with an information sheet that explained the pilot and informed them that their participation would involve the possibility of being assigned to the blended section.[12] Prior to randomization, consenting students completed a baseline survey that asked questions about demographic and socioeconomic background, as well as information on students' interest in the course, access to technology, and experience taking online courses previously.

A total of 827 students were recruited to participate in one the study's 7 courses, and 700 ultimately consented.[13] The majority of participating students were recruited and assigned to either the treatment or control group after the first week of the semester. A second

[11] Despite efforts made to reduce access to content among control students, course endline surveys reveal some minimal cross over. I discuss implications in Section 5.1.

[12] A limitation of the study design is that students declining to participate in the study attended the traditional lecture sections. As a result, the traditional sections were larger in class size (as shown in Table 1) and had arguably differing peer effects (containing students not electing to participate in the study). I address this limitation in Section 3.5.3.

[13] The sample of 700 student observations includes multiple observations of students who enrolled in more than one of the courses in the study – 34 students enrolled in two courses, and 3 students enrolled in 3 courses. I treat students in multiple courses as separate observations as learning outcomes are course-specific, and treatment students were only given access to the courses in which they were assigned to treatment.

round of randomization was conducted after the second week of the semester for the handful of students who enrolled in the courses late.[14] After each week of recruitment, consenting students were randomly assigned into either the treatment (blended format) or control group (traditional format). The randomization was stratified by week of enrollment, course, semester, instructor, and gender and split students roughly evenly into treatment and control groups.

3 Data and Estimation Strategy

3.1 Data

A. Student Surveys

At baseline, students provided information on demographic and socioeconomic background, interest in the course, access to technology, and previous experience with online learning. At the end of the semester, students completed an endline survey that probed measures of participation, engagement with course material, time spent on coursework, interest in future courses, and course satisfaction. Students completed the endline survey during the week of their final exam but prior to receiving their grades on the final exam or course.

B. Final Course Grades

For each of the courses, the university registrar provided the final course scores of students participating in the study. Course scores are measured on a 100-point scale and assigned at the discretion of course instructors. The registrar received the scores directly from instructors at the end of each semester. The field team requested instructors to provide

[14] 41 students (across 5 courses in semester 1 and 1 course in semester 2) enrolled in the second week.

exam scores and the final course score; however not all did. Hence, the registrar-provided course score is the best available measure of student learning in the course.[15]

C. Transcript Data

The university registrar also provided full transcript data (i.e., a record of every course in which participating students enrolled and the grade assigned in each), two years after the second semester of the study concluded. These data allow me to control for pretreatment GPA and to examine the impact of treatment on subsequent course completion and performance.

D. Online Platform Analytics

The university provided data analytics collected through the learning management system on students' activity on the online platform. Specifically, the university shared the number of times students viewed each weekly video and the proportion of each video watched.

E. Classroom Observations

The field team conducted one classroom observation for each section of each course-professor-semester combination in weeks 10-11 of each semester.[16] Observers collected

---

[15] Instructors were asked to provide data on final grade, as well as student attendance and on the outcomes of quizzes and exams. However, professors shared data inconsistently across courses and it is therefore difficult to make meaningful comparisons. In the first semester, two of the instructors failed to provide attendance and quiz data and four failed to provide final exam scores. In the second semester, one instructor neglected to provide attendance and quiz data and two failed to provide final exam scores. 578 professors provided students' final grades, which I compared with those supplied by the registrar. The registrar's reported scores generally matched the scores provided by instructors (see Appendix Figure A2 and Section 4.1 for a broader discussion of the source of course scores). For roughly 79 percent of students (n=551), instructor-provided grades were confirmed to be identical to registrar-provided grades. 9 percent of students' grades (n=63) were reported by the registrar but not by the instructor. 3 percent of students' grades (n=21) were reported by the instructor but not by the registrar. Roughly 9 percent of the students (n=65) had grades that were reported differently by the registrar and instructor (only n=17 of the students in this category received a passing grade). The average difference between the grades reported by the registrar and instructor in the latter category was 10.7 score points.

[16] In total, 29 observations were conducted over the 14 course-professor-semester offerings. For two courses, professors held two control lecture sections due to large class size, and in one of these courses, the professor also held two treatment discussion sections. For two courses, observers were unable to observe a treatment

information on instructor and student attendance and engagement and took recorded the time spent on different class activities. These observations were intended to illuminate how instructors used their section time in the blended format as opposed to the traditional format.

F. Qualitative Instructor Interviews

The field team also conducted open-ended interviews, roughly 1 hour in length, with each of the instructors upon the conclusion of each semester. Interviewers followed a semi-structured interview protocol which inquired about instructors' experiences making and teaching online content and the contrasts with the traditional lecture style used in the study years and in previous years. Interviews were audio recorded and transcribed and translated into English for analysis.

3.2 Integrity of Experimental Design

Table 2 shows balance between treatment assignment pretreatment characteristics collected from the baseline survey. The adjusted differences control for strata dummies, and robust standard errors are used. Across the 29 characteristics examined, one difference is statistically significant at the 5 percent level. A joint test of significance of pretreatment characteristics, regressing treatment assignment on all characteristics included in Table 2 (and controlling for strata dummies and using robust standard errors), is not significant (p= 0.61).[17] Overall, these tests point to overall balance and a successful randomization. Following accepted practice in the experimental literature (Duflo et al., 2007; Bruhn & McKenzie, 2009), I control for pretreatment characteristics in my treatment effect estimation; however, results are robust to the inclusion and exclusion of these control variables.

section because the professors were not holding regularly holding in-person discussion sections. Treatment and control observations were conducted in the same week for each course.

[17] This joint test is limited to observations for which there are no missing data across characteristics (N=404).

[Table 2]

As shown in Table 2, the majority of students appear to have access to computers and internet at home. Around 90 percent of students took the course included in the study because of a degree requirement. Around 30 percent were re-taking the course after previously received a failing grade for the prior semester enrolled. Students primarily major in subjects related to electronics, information technology and computing.

I also examine two main sources of attrition: (1) missing a final course score and (2) missing endline survey data (see Appendix Table A1). With regard to the first, roughly 6 percent of students are missing a final course score. These individuals attended the course during the first two weeks of the semester (and hence were recruited for the study), but never officially enrolled in the course, (and hence have no record of being enrolled with the registrar). The missing rate is the same across treatment and control, and thus it does not appear that assignment to treatment affected students' enrollment decisions.

With respect to the endline survey, roughly 19 percent of students were not found or did not participate in the endline survey. The attrition rate of the control group (22 percent) is somewhat higher than the treatment group (16 percent), with differences driven by those not found rather than declining participation (see Appendix Table A1).[18]

3.4 Empirical Strategy

I estimate the average effect of assignment to the blended format through the following intent-to-treat estimation strategy:

$$Y_{ic} = \beta_0 + \beta_1 \, treat_{ic} + \delta_c + \mathbf{X}'_i + \varepsilon_{ic}$$

[18] While the differential attrition was minimal, I discuss implications for comparisons of engagement and satisfaction in Section 5.2.

where Y represents the outcome of interest of student *i* in course-semester-professor combination, *c*, and *treat* is a dummy for whether student is assigned to a blended online treatment section. β_0 is a constant. β_1 is the treatment coefficient of interest and will reflect differences in outcome for the blended treatment sections. I also run the specification including randomizations strata fixed effects (δ_c), as well as a vector of student pretreatment covariates ($\mathbf{X}'_i$). ε_{ic} represents the error term. I use ordinary least squares (OLS) for continuous outcomes and binary outcomes. Because the number of clusters (i.e. face-to-face sections held by professors) is small (N=30), I run specifications using robust standard errors as they are more conservative than clustering standard errors by course-semester-professor sections.

4 Main Results

4.1 Distribution of Course Scores

An examination of course scores reveals a very high failing rate among both treatment and control students. Students must receive a score of 60 or higher to pass the course and receive credit. Figure 1 displays the cumulative distribution of course scores.[19] Roughly 46 percent of students received a score less than 60, and 19 percent of students received a score of zero.[20]

[Figure 1]

A score of zero comes from two sources. (1) Zeros are directly assigned by professors to reflect that the student completed little to none of the assigned coursework. (2) Zeros are

[19] The university grading categories are the following: F (0-59), D- (60-64), D+ (65-69), C- (70-74), C+ (75-79), B- (80-84), B+ (85-89), A- (90-94), A+ (95-100).
[20] University administrators confirmed that the high failing rate among the courses was similar to previous years, with roughly 40-50 percent of students failing courses.

assigned by the registrar when a student officially withdraws from a course, the registrar records the official course score as zero. Because the registrar only provided me with course grades on the 100-point scale (and no letter grades reflecting official withdraws), I am unable to determine the source of the zeros within my sample. However, an examination of the cases where I have both registrar- and professor-assigned grades suggests that a minority of the zeros reflect an official withdrawal.[21]

4.2 Impact on Withdrawal and Failing Rate

Table 3 presents the results of the main ITT specification on the probability of receiving a score of zero (i.e., withdrawal and/or disengagement from the course) and the probability of receiving a score of less than 60 (i.e., failing). For the former, I find that assignment to treatment increases the probability of withdrawal/disengagement (receiving a zero) by 5-6 percentage points (significant at the 5 percent level when controlling for pre-treatment GPA). I find no significant difference between treatment and control – overall the probability of failing the course (receiving less than 60 points) is the same for treatment and control. The failing students includes the students receiving zeros; hence, the treatment impact inducing withdrawing or disengaging from the course is not impacting the overall passing rate.

[Table 3]

The finding that students in the treatment group are more likely to have a score of zero may reflect an initial resistance to the new blended online format, causing them to officially withdraw or disengage completely from the course. Figure 2 shows the proportion of treatment students viewing each week's videos across the 16-week semester, disaggregated

[21] Of the 578 cases for which I have both registrar and professor-assigned grades, 103 students have a registrar grade of zero. Of these 103 students, 26 students (across 5 courses) have professor-assigned grades that are greater than zero. All of the professor-assigned scores are less than 60, and the average across these 26 cases is 10.9 points. Having a mismatched score is balanced across treatment and control. The full distribution of registrar-assigned and professor-assigned scores is shown in Appendix Figure 2.

by course score. Across all treatment students, viewership is not high and declines as the semester progresses. Viewership is lowest among students scoring zero (only at 20 percent in the first two weeks and dropping to below 10 percent by week four). Appendix Table A2 likewise shows low rates of engagement with course activities among students scoring zero. While I am unable to observe the timing at which students decide to withdraw or disengage from the course, the platform analytics suggest that this decision is made early on.

[Figure 2]

4.3 Impact on Course Score

Table 4 presents the results of the main ITT specification on overall course score (raw score and standardized scores).[22] In the sparest specification that excludes baseline covariates, I find a small negative, but statistically insignificant effect on overall course score. I use Akaike's Information Criterion (AIC) to determine the best model fit which privileges my specification in column 3 with pretreatment GPA and strata fixed effects. After controlling for pretreatment GPA and strata fixed effects, I find that the students assigned to treatment score 3 course points or 0.06 standard deviations lower than the control group. Both estimates are of similar magnitudes and the inclusion of pretreatment GPA results in increased precision as evidenced by the reduction in the standard error of 7 percent.

[Table 4]

As a robustness check given the differential withdrawal among the treatment group, I winsorize regressions at the 25th, 50th, 75th, and 90th percentiles to remove variation from the lower end of the distribution and examine the effects for students in the upper end of the score distribution. Specifically, I replace scores of zero and scores below the percentile of

[22] Scores are standardized within course-professor-semester grouping.

interest with the score at that percentile. I use the control distribution to determine the percentile cut point scores. The winsorized regressions also show a small negative but statistically insignificant effect for students in the upper end of the distribution. Overall, it appears that treatment assignment does not significantly impact student scores.

As noted previously, student course score is the best available measure of student learning; however, there are reasons why course scores may not accurately reflect student learning. Heaping at letter grade thresholds (e.g., at 60, 70, and 80 points, as can be observed in Appendix Figure 2) suggests that instructors might be inclined to inflate grades in certain circumstances. While non-differential grade inflation could limit the ability of the course score to objectively reflect the skill level and learning of the student, differential grade inflation between treatment and control would pose a threat to internal validity. As the pilot was an initiative led by course instructors (i.e., course instructors created online videos and are motivated to use them in subsequent years), instructors might have been inclined to bias grading in favor of the treatment group. However, evidence examined does not suggest that differential grading occurred. I find no significant difference on the likelihood of failing, and the treatment impact on overall course score is negative, suggesting that control students might perform slightly better. Furthermore, checks at the threshold grades suggest there is no differential bias toward inflating treatment students' grades toward a higher letter grade (see Appendix Table A3).

4.4 Heterogeneity

There are reasons to believe that the treatment impact might differ for students of varying ability. Students of higher ability may possess skills important for adjusting to a new system of learning. They may likewise be more capable of self-regulating the arguably independent learning approach of the blended model.

A handful of studies suggest that success in online learning may depend on students' ability to manage time and self-direct their learning (Banerjee & Duflo, 2014; Michinov et al 2011; Lu et al 2003; Xu and Jaggers, 2013; Stewart et al 2010). Donovan, Figlio and Rush (2006) find that cramming is pervasive among students completing online courses. Because the blended model's success depends largely on the students' ability to optimally view online lectures, when students fail to regulate learning, this model may be less effective. Indeed, Figlio et al. (2013) find strong negative effects on achievement among male, Hispanic and lower-achieving students, suggesting that the use of online courses may be particularly detrimental for disadvantaged students.

I use pretreatment GPA prior to each student's enrollment as a measure of ability prior to the intervention and disaggregated students by quintiles.[23] As shown in Table 5, the treatment impact on increasing the likelihood of receiving a zero does not extend to students in the upper quintile. Not surprisingly, students in the top quintile of prior GPA in both treatment and control are less likely to have a score of zero, but top quintile students assigned to treatment are also no more likely to receive a score of zero. This suggests that the initial resistance to assignment to the blended model inducing students to withdraw or disengage is not happening among the highest ability students.

[Table 5]

4.5 Impact on longer term academic outcomes

As shown in Table 6, I find that assignment to the blended model does not impact students' ultimate trajectory. I use student transcript data for the two years post intervention to examine impacts of the treatment assignment on longer-term outcomes. There are no significant differences between treatment and control in the rate at which students left the

[23] I also disaggregated students into quartiles and terciles and find substantively similar results.

program without obtaining a degree and no difference in the number of credits taken and students' GPA post treatment.

[Table 6]

Following the study, the university transitioned to a hybrid or supplemental version of the flipped classroom model in which all faculty continued to use and provide online video content to students to view outside class time, but decided *not* to maintain the shorter in-person sessions of 50 minutes in duration. Rather, in-person sessions continued as in years prior to the study at 90 minutes and faculty had discretion to use the 90-minute sessions as they wished. Faculty reported using the sessions both to introduce lecture material and to review material provided previously through lecture videos.

Hence, although assignment to the blended model may have induced students to withdraw or disengage (i.e. earn a score of zero) at a greater rate in the short run, ultimately there is no difference in their overall completion rate of the focal courses (i.e. the courses in which they were enrolled for this study). Across both treatment and control, roughly 71 percent passed the focal course two years post treatment. Overall, roughly 22 percent of students retook the focal course, and treatment students were 5 percentage points more likely to retake the course (marginally significant at the 10 percent level).[24] Among treatment students receiving a zero, 52 percent retook the focal course, of which 44 percent of retaking passed upon subsequent attempt (proportions not shown in table). While it is not clear that having the opportunity to retake the course with the 90-minute in-person sessions may have helped with re-take passing, it is possible that this opportunity benefitted some of the students who initially withdrew as a result of assignment to treatment.

[24] Among the students retaking the focal course, 42 percent received a score of zero during the study and 98 percent receive a score less than 60. Among the students failing the focal course during the study, 47 percent retook the course.

5 Compliance and Course Experience

The findings of this study show there is little difference in student outcomes for those assigned to the blended model compared to those assigned to the traditional format. While this suggests that the blended model is as effective as the traditional format, imperfect compliance with treatment assignment might contribute to the similar outcomes. In this section, I explore (1) the extent to which students complied with assignment as well as (2) compare the course experiences between blended and traditional formats.

5.1 Compliance

Course instructors and department administrators made efforts to reduce the prevalence of noncompliance by discouraging students from attending sections to which they were not assigned and by not providing control students login access to the online course platform. Nevertheless, qualitative interviews with instructors suggest that control students were able to access online video content, likely through peers. Unfortunately, I am unable to determine whether treatment students attended traditional lectures due to insufficient attendance records, but this remains a possibility. The main concern is that students in the treatment and control group were able to change their condition and/or access course content through a preferred method. The direction of bias is ambiguous and would attenuate the main results unless students in one condition were more likely eschew their assigned condition. In this section I explore the two ways non-compliance may influence the results.

First, I check whether control students had access to the treatment condition. As shown in Table 7, none of the control students were able to officially log onto the online course platform. However, in the endline student surveys, 8 percent of control students said they were provided access. 23 percent of control students reported they were able to access course videos (even if access was not given), and 15 percent reported that they watch course videos in a typical week.[25] At the same time, the average number of minutes reported watched in a typical week by the control is significantly lower than that reported by the treatment group (only 11.18 minutes compared to 107.71 minutes on average among treatment students).

[Table 7]

Second, I check whether treatment students accessed the control condition. Unfortunately due to poor attendance records, it is less clear whether treatment students (particularly those that received passing scores) opted to skip watching videos and attended traditional lecture sections instead.[26] As shown in Figure 2 and Table 7, video viewership among the treatment is not high, even among students who pass the class. Furthermore, 48 percent of students that did not log on to the platform received a passing grade. It may be the case that these students used textbooks or other sources of information to pass the class. For most of the classes, grades were based on performance on exams and quizzes. Attendance and/or video viewership did not officially factor into instructors' grading rubrics. In follow-up qualitative interviews, instructors noted that attendance has always been low in previous years (with roughly 40-50 percent of students attending regularly); hence the lower video viewership could be a reflection of this type of course-taking behavior as well rather than treatment students attending control lectures in lieu of viewing videos.

While the type of noncompliance observed would likely bias results toward zero, there are reasons to believe this might be minimal. For one, the intensity of video viewership among control students does not appear high. Second, it is arguable that the students who

25 Students were asked to report the number of minutes they spend on a series of course-related activities, including whether they watched official course videos. Treatment and control students were asked to fill out identical surveys. Hence, we did not ask control students how they access the videos in order to not bias their responses such that they withheld information about accessing treatment content. Qualitative interviews with instructors revealed that some treatment students logged on with classmates to watch videos (and thus would not appear in clickstream data as having watched videos).

26 Instructors were asked to provide attendance records for treatment and control sections, but the method of attendance taking and accuracy differed across courses. For a subsequent draft, I plan to investigate more fully to gain a better sense of whether this occurred.

would go to lengths to obtain access to videos are not at the margin where I would see the most movement in my results. They are likely highly motivated and would be more likely to perform well in courses regardless of video access. Indeed, the majority of control students who reported being able to access videos and who reported watching videos were in the upper three quintiles of the pre-treatment GPA distribution.

5.2 Course Experience

Endline student surveys, as well as classroom observations conducted in the latter half of each semester provide a more comprehensive understanding of the treatment contrast experienced by students in the study. In this section, I examine (1) fidelity to the "flipped-classroom" design of the blended treatment and (2) student engagement and satisfaction.

A. Fidelity to Blended Course Design

With respect to fidelity to the treatment design, evidence suggests that students did not adhere to a strict interpretation of the blended design. In Panel C of Table 7, I show that students in the treatment were significantly less likely to attend the face-to-face section in a typical week. On average, they reported attending fewer weeks and felt that in-person sections were less useful than the control group. In fact, nearly a quarter of treatment students reported never attending face-to-face sessions, but were active on the online platform and hence their course experience was fully online rather than blended.[27]

The lower levels of participation in face-to-face sessions among treatment may have been because students felt video lectures were sufficient substitutes for in-person time; however, it may also have been because instructors themselves were treating videos as

---

[27] 24 percent of treatment students enrolled in the online platform, but reported not attending any face-to-face sessions. Similarly, however, 24 percent of treatment students reported attending face-to-face sessions, but never enrolled in the online platform.

sufficient substitutes for in-person instruction.[28] However, the direction of causality cannot be established – instructors may have ceased holding in-person sessions due to lack of student interest, which was reflected in qualitative interviews with instructors. The differences in the way online models were carried out provide, in theory, an opportunity to examine treatment heterogeneity by whether the instructor treats the online model as a substitute to a face-to-face instructor or a complement. I conducted this analysis, but find no heterogeneity in treatment impact – this could be due to the small size of the two courses in which face-to-face sections were held. The question of whether students' interest drove this manifestation of the model also confounds the analysis.

At the same time, nearly a quarter of treatment students never enrolled in the online platform, but reported attending the in-person sessions, resulting in a more inferior fully face-to-face course experience as these treatment sessions were never meant to replace the delivery of lecture material. Classroom observations suggest that the treatment in-person sections were largely run as "flipped-classroom" style discussion sections. As shown in Table 8, none of the observations were instructors observed teaching new material, and instructors confirmed in qualitative interviews that they did not repeat lecture material in discussion sections. Rather, sessions were used to review prior week material, Q&A, and reviewing the online videos.

[Table 8]

The diluted nature of the treatment design suggests that under a stricter adherence to a flipped model, treatment students may have performed better than as observed. However, given the independent nature of higher education studies, the take-up of video versus face-to-

---

[28] For two courses, observers were unable to conduct a classroom observation because they discovered that instructors were not regularly holding face-to-face sessions for treatment students. Indeed, in these two courses, less than 50 percent of students reported attending the in-person session or finding it useful.

face sessions is in itself an interesting finding about student choices in more self-directed blended online learning environments.

B. Student Engagement and Satisfaction

Overall, I find student engagement and satisfaction, as reported in the student surveys, to be equivalent across treatment and control. As shown in Table 7, the time spent on activities in a typical week (e.g. meeting with a professor 1-on-1, time spent studying, etc.) is largely the same across both groups. Satisfaction is likewise similar (with the exception that treatment students are less likely to find the in-person section to be useful, as discussed above). As shown in Table 8, the classroom observations suggest that engagement might be higher in treatment sections among treatment students that actually attended – observer assessments find a higher proportion of control observations in which more than half of students were unengaged (i.e. showing signs of boredom and not interacting with peers or the instructor).

As noted in Section 3.2, for the endline survey, the attrition rate of the control group (roughly 22 percent) is higher than the treatment group (roughly 16 percent), with differences driven by those not found rather than declining participation. Because the field team first attempted to locate students for the endline survey at the in-person sections, students found at endline might be more inclined to report more engagement and/or higher satisfaction with the course and with in-person sections. If control students who did not attend in-person sections experienced lower levels of satisfaction than those who did, then the true difference in satisfaction with the in-person sections among treatment students might would be attenuated compared to the observed findings.

6 Discussion

In recent decades, the advancement of computing technologies has resulted in a proliferation of new educational resources, including efforts to bring classrooms into online settings. This wave in online learning has reached lower income countries where governments and donors are increasingly looking to online courses to help education providers lower costs and improve access to high quality teaching content. However, little rigorous evidence exists about whether online and distance instruction is effective these settings where student needs may differ considerably from those in higher income nations, where most research in this area is conducted.

This study provides some of the first experimental evidence on the effectiveness of online learning in a lower-middle income country by evaluating a blended pilot in STEM courses at a public university in Mongolia. While it appears that assignment to the online model leads to initial resistance (i.e., a higher likelihood of withdrawing or disengaging from courses) among lower ability students, performance in the courses (both passing rate and overall course score) was comparable between those in the online model compared to the traditional face-to-face model. The comparable performance suggests that a blended online may be nearly, if not equally, effective in producing the same amount of student learning for most students. In the long run, student academic outcomes are also equivalent across treatment and control groups. However, these long-run outcomes might reflect that some of the initially withdrawing treatment students may have benefitted from the opportunity to retake the course with longer in-person instructional time, as was instituted by the university after the study.

While experimental studies of blended learning models have found similar results – that blended models are as effective as traditional in-person teaching – the setting of this study differs dramatically from the settings in which online models have been experimentally or quasi-experimentally evaluated (Alpert et al., 2016; Bowen et al., 2014), and even in settings in the U.S., findings suggest lower ability students fare worse in online settings (Bettinger and Loeb, 2017). In the Mongolian context, course absenteeism and failing rates are extremely high, which may reflect lower student capacity and the need for more in-person instructional support.

The open-ended interviews with course instructors reflected the challenges of the teaching setting and the acclimation that was needed on the student side to embrace the new model, with multiple instructors noting that their treatment students needed at least 3-4 weeks to adapt to the online model. Additionally, the discussion-style format of the treatment sections was not a norm in the academic department, and instructors both had to learn how to ask meaningful questions to prompt student participation, while students took time to figure out how to interact and engage with instructors in a beneficial way. Nearly all instructors emphasized that the in-person sections were important for keeping students on track, while also acknowledging that not much could be done about the low attendance in either treatment or control settings. Heterogeneity analyses confirm that the higher likelihood of withdrawal was concentrated among lower performing students, who may need more scaffolding in new learning environments.

While the department decided to maintain the longer in-person sessions, in interviews, instructors' collective assessment of the model confirmed the overall quantitative findings – that student performance was roughly equivalent between treatment and control for most students. The majority of instructors also expressed that the shortened 50-minute in-person sections were sufficient, particularly in light of efficiency gains with respect to their time and the ability to redirect efforts to other tasks. These gains in faculty time may be particularly important in lower-resource settings where there are fewer content-matter experts available, not just for instructing students, but also for conducting research that is valuable for countries' economic growth and development. Ultimately, if student learning is roughly equivalent, a move toward blended online models may prove to be a more pareto efficient option for institutions.

The availability of the video lecture content also offers more flexibility and efficiency for students, who have another option beyond physically attending lectures. Student surveys revealed that roughly a quarter of treatment students only viewed lecture videos and did not attend in-person sections. Even so, performance among treatment students in the course was similar to that of the control group. If students can appropriately select into using the course materials most appropriate for their learning, the model may also have added benefits for students with regard to provide more flexible learning options.

The success of the blended model might also allow institutions to improve access to content by increasing class size, particularly if a substantial proportion of students self-select into a fully online version of the course (i.e., only watch lecture videos). Increasing online class sizes does not result in substantial increases to operational cost, and small increases in class size may not be detrimental to student learning in online settings (Bettinger et al., 2017). The generalizability of this pilot study to larger class sizes is limited because students were split into two concurrently-run sections (and hence the in-person class sizes were considerably smaller than in a scale-up in which only one in-person section is held); however, the longer-run analysis of pass rates suggests equivalent pass rates between treatment and control, even with the institutional move to a hybrid of the flipped model.

Although this study cannot ultimately speak to the implications for improvements to access to teaching content outside a blended model, the success of the model sets the stage for future work to examine how fully online models can improve learning in areas where there is no high-quality face-to-face instruction. For instance, teaching models which couple high-quality content with untrained or minimally trained teaching assistants have been shown to successfully raise learning in low resource settings (Banerjee et al., 2007). The digital content generation coming out of efforts to transition to online learning may also be particularly important for lower income countries, where there is a shortage of content made in local languages and targeting local populations. For instance, the university has made two of the online courses coming out of this pilot into MOOCs available for public access. More work is needed to understand whether centrally created content can be successfully disseminated to satellite campuses and/or publicly to improve learning either through a fully online model or through a blended model utilizing teaching assistants.

References

Acemoglu, D., Laibson, D., & List, J. A. (2014). Equalizing superstars: The Internet and the democratization of education. *The American Economic Review*, 104(5), 523-527.

Alpert, W. T., Couch, K. A., & Harmon, O. R. (2016). A randomized assessment of online learning. *The American Economic Review, 106*(5), 378-382.

Altbach, P. G. (2009). Peripheries and centers: Research universities in developing countries. *Asia Pacific Education Review*, 10(1), 15-27.

Banerjee, A., Cole, S., Duflo, E., & Linden, L. (2007). “Remedying education: Evidence from two randomized experiments in India.” *Quarterly Journal of Economics* 122 (3): 1235–1265.

Banerjee, Abhijit V., and Esther Duflo. (2014). “(Dis)organization and Success in an Economics MOOC.” *American Economic Review 104*(5): 514–18.

Barro, R. J., & Lee, J. W. (1993). International comparisons of educational attainment. *Journal of monetary economics*, 32(3), 363-394.

Barro, R. J., & Lee, J. W. (1996). International measures of schooling years and schooling quality. *The American Economic Review*, 86(2), 218-223.

Barro, R. J., & Lee, J. W. (2001). International data on educational attainment: updates and implications. *Oxford Economic Papers, 53(3), 541-563.*

Bettinger, E., Fox, L., Loeb, S., & Taylor, E. S. (Forthcoming). Virtual Classrooms: How Online College Courses Affect Student Success. *The American Economic Review*.

Bettinger, E., Doss, C., Loeb, S., Rogers, A., & Taylor, E. (2017). The Effects of Class Size in Online College Courses: Experimental Evidence. *Economics of Education Review*.

Bettinger, E., & Loeb, S. (2017). Promises and Pitfalls of Online Education. *Evidence Speaks Reports*, 2(15). Center on Children and Families at Brookings.

Bowen, W., Chingos, M., Lack, K., & Nygren, T. (2014). “Interactive learning online at public universities: Evidence from a six-campus randomized trial.” *Journal of Policy Analysis and Management* 33 (1): 94–111.

Bulman, G., & Fairlie, R. W. (2016). *Technology and education: Computers, software, and the internet* (No. w22237). National Bureau of Economic Research.

Bruhn, M., & McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American economic journal: applied economics*, 1(4), 200-232.

Chaudhary, L., & Bacolod, M. (2016). Distance to Promotion: Evidence from Military Graduate Education. Calhoun International Archive of the Naval Postgraduate School.

Cheney, C. (2017, April 3). The road to real results for online learning in developing countries. *Devex Development Newswire*. Retrieved from https://www.devex.com/news/the-road-to-real-results-for-online-learning-in-developing-countries-89884

Cowen, T., & Tabarrok, A. (2014). The industrial organization of online education. The American Economic Review, 104(5), 519-522.

Deming, D. J., Goldin, C., Katz, L. F., & Yuchtman, N. (2015). Can Online Learning Bend the Higher Education Cost Curve? *The American Economic Review,* 105(5), 496-501.

Donovan, C., Figlio, D. N., & Rush, M. (2006). *Cramming: The effects of school accountability on college-bound students* (No. w12628). National Bureau of Economic Research Working Paper. Retrieved from http://www.nber.org/papers/w12628

Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: A toolkit. *Handbook of development economics, 4*, 3895-3962.

Figlio, D, Rish, M., & Yin, L. 2013. "Is It Live or Is It Internet? Experimental Estimates of the Effects of Online Instruction on Student Learning." *Journal of Labor Economics* 31(4): 763-784.

Garrido, M., Koepke, L., Andersen, S., Mena, A., Macapagal, M., & Dalvit, L. (2016). *An examination of MOOC usage for professional workforce development outcomes in Colombia, the Philippines, & South Africa.* Seattle: Technology & Social Change Group, University of Washington Information School.

Goodman, J., Melkers, J., & Pallais, A. (2016). *Can Online Delivery Increase Access to Education?* (No. w22754). National Bureau of Economic Research.

Hart, C. M., Friedmann, E., & Hill, M. (2016). Online course-taking and student outcomes in California community colleges. *Education Finance and Policy.*

He, F., Linden, L., & MacLeod, M. (2008). *How to Teach English in India: Testing the Relative Productivity of Instruction Methods within the Pratham English Language Education Program.* Columbia University: Working Paper.

Joyce, T., Crockett, S., Jaeger, D. A., Altindag, O., Center, C. G., & O'Connell, S. D. (2014). *Does Classroom Time Matter?* (No. w20006). National Bureau of Economic Research Working Paper.

Kizilcec, R. F., Saltarelli, A. J., Reich, J., & Cohen, G. L. (2017). Closing global achievement gaps in MOOCs. *Science*, 355(6322), 251-252.

Kizilcec, R. F., & Halawa, S. (2015). Attrition and achievement gaps in online learning. *In Proceedings of the Second (2015) ACM Conference on Learning@ Scale (pp. 57-66).* ACM.

Lai, F., Luo, R., Zhang, L., Huang, X., & Rozelle, S. (2015). Does computer-assisted learning improve learning outcomes? Evidence from a randomized experiment in migrant schools in Beijing. *Economics of Education Review*, 47, 34-48.

Liyanagunawardena, T., Williams, S., & Adams, A. (2013). The Impact and Reach of MOOCs: A Developing Countries' Perspective. eLearning Papers.

Lovett, M., Meyer, O. & Thille, C., (2008). JIME - The Open Learning Initiative: Measuring the Effectiveness of the OLI Statistics Course in Accelerating Student Learning. *Journal of Interactive Media in Education. 2008*(1), p.Art. 13. Retrieved from http://doi.org/10.5334/2008-14

Lu, J., Yu, C., & Liu, C. (2003). Learning style, learning patterns, and learning performance in a WebCT-based MIS course. Information & Management, 40:497-507.

Means, B., Toyama, Y., Murphy, R., Bakia, M., and Jones, K. (2010) *Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies.* U.S. Department of Education, Office of Planning, Evaluation, and Policy Development. U.S. Department of Education. (2012) Understanding the Implications of Online Learning for Educational Productivity. Office of Educational Technology, Washington, D.C.

Michinov, N, Brunot, S., Le Bohec, O., Juhel, J., & Delaval, M. (2011). Procrastination, participation, and performance in online learning environments. Computers and Education, 56:243-252

Sodnomtseren, A. (2012). Financing Public Higher Education Institutions in Mongolia.

Stewart, C., Bachman, C., & Johnson, R. (2010). Students' characteristics and motivation orientations for online and traditional degree programs. *Journal of Online Learning and Teaching*, *6*(2), 367−379.

Streich, F. (2014). "Online Instruction and Student Success in College: Evidence from Community Colleges." *Online Education in Community Colleges: Access, School Success, and Labor-Market Outcomes.* Dissertation submitted in partial fulfilment of requirements for Doctor of Philosophy, University of Michigan.

Robertson, A. (2015, February 11). Can online classrooms help the developing world catch up? *The Verge*. Retrieved from https://www.theverge.com/2015/2/11/8014563/bill-gates-education-future-of-online-courses-third-world.

UNESCO Institute for Statistics. (2012). Data Centre. Retrieved from http://stats.uis.unesco.org

Xu, D., & Jaggars, S.S. (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*(3), 360−377.

Xu, D., & Jaggars, S. S. (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.

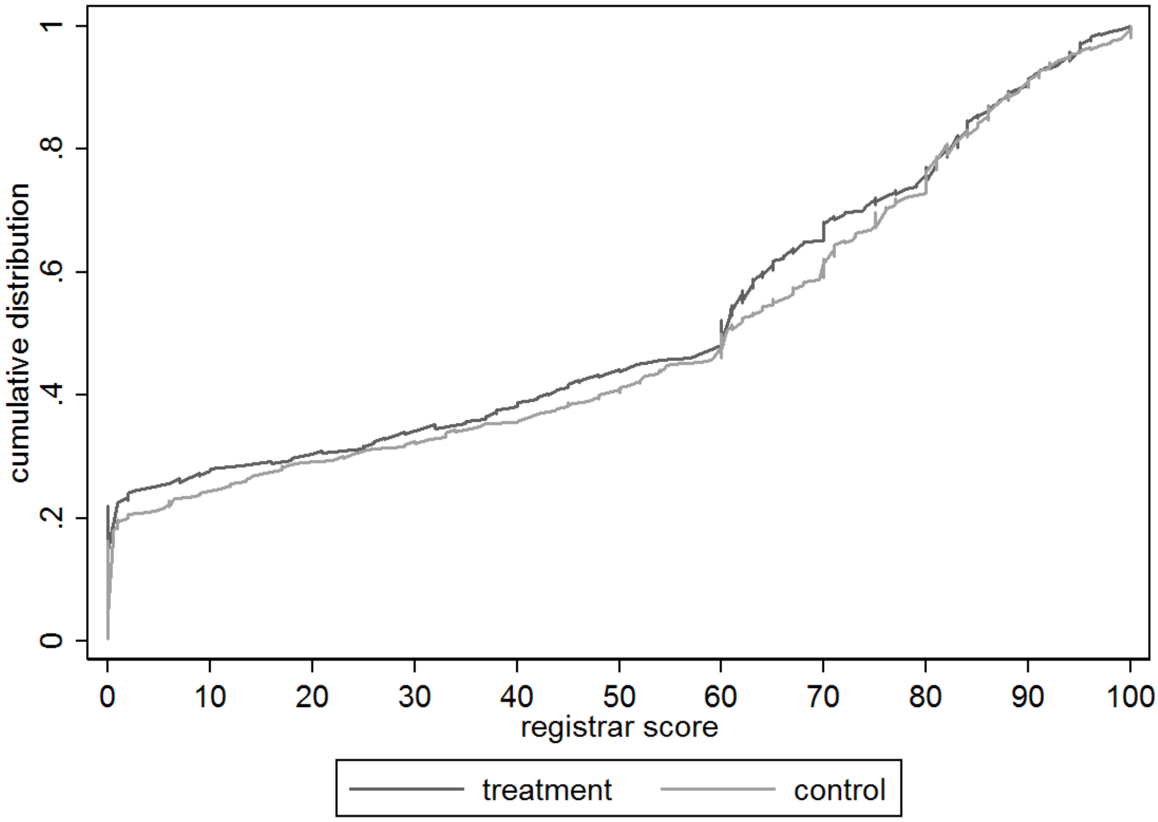


Figure 1. Cumulative distribution of final course score, by treatment assignment

Notes: This figure shows the cumulative distribution of final course scores provided by the university's registrar, by assignment to the blended treatment.

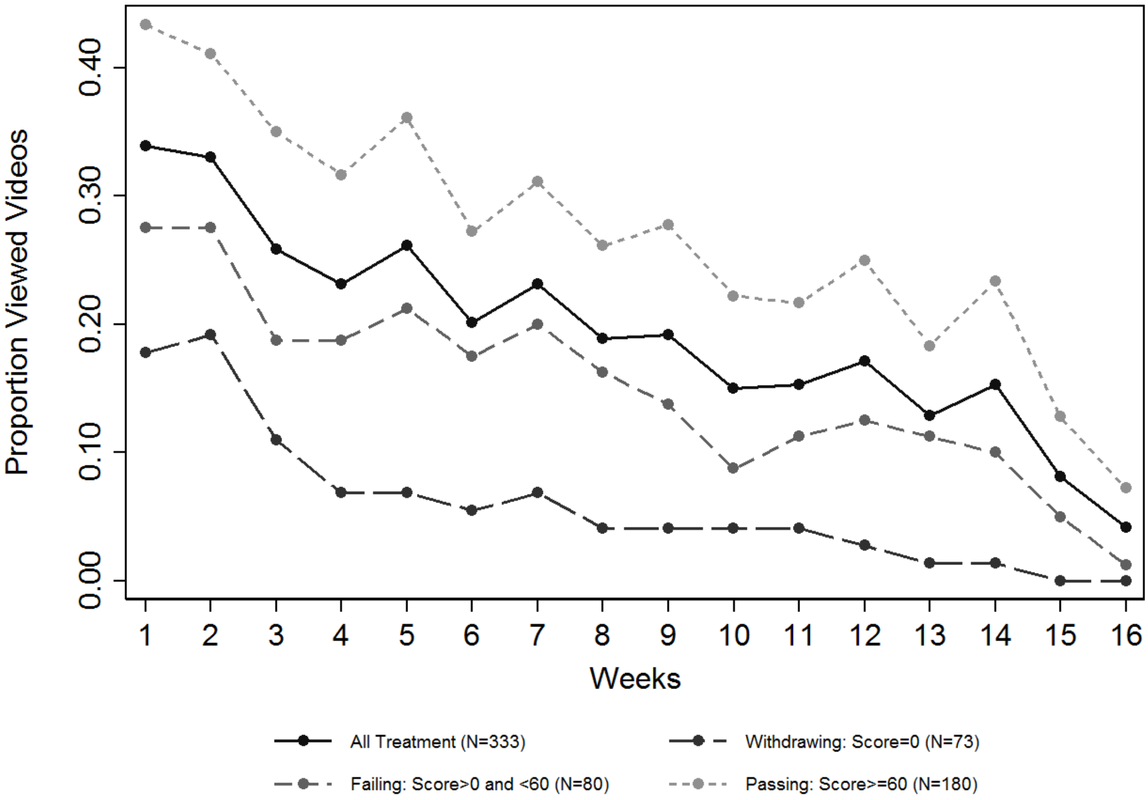
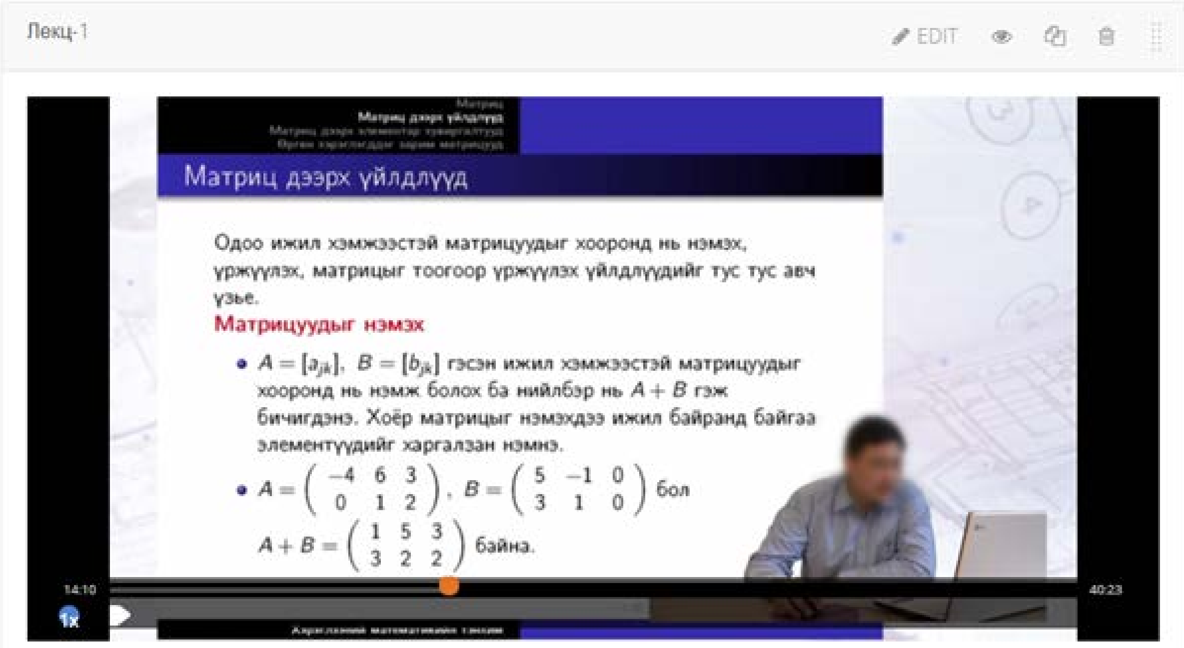


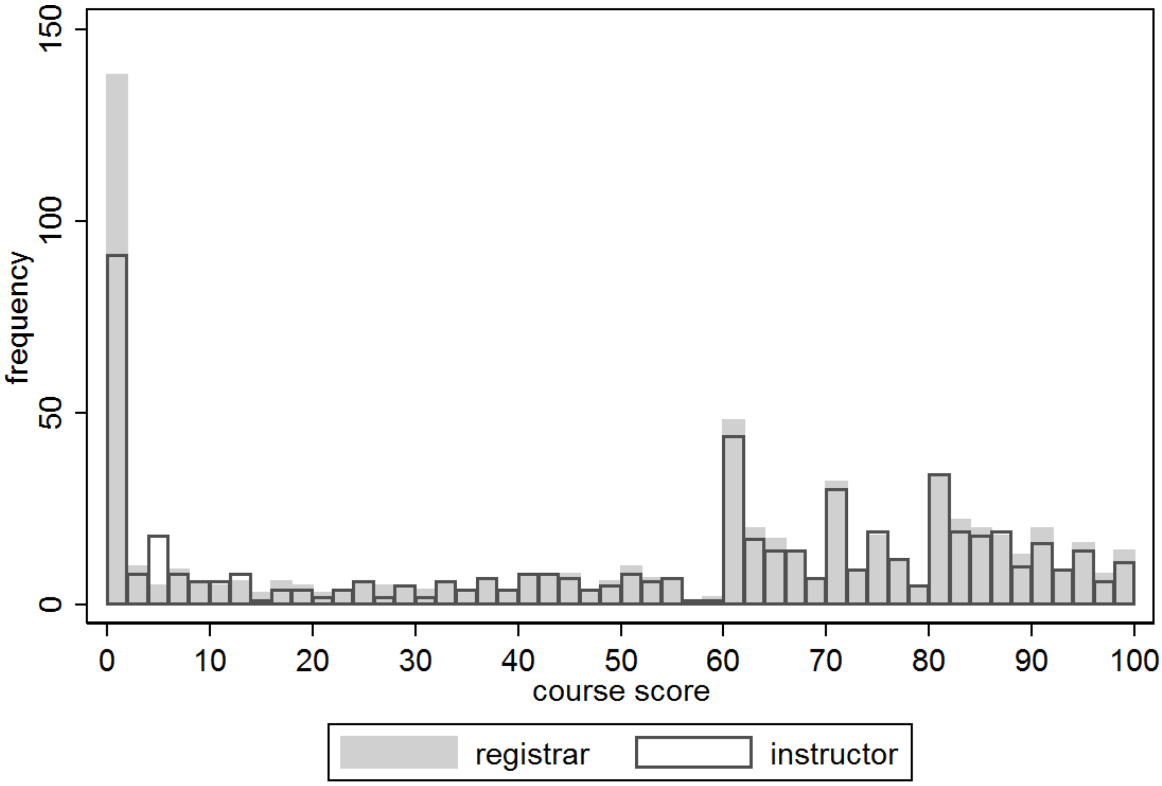
Figure 2. Proportion of students viewing videos, by week and course score

Notes: This figure shows the proportion of students viewing the videos associated with each week's content. Students were defined as having viewed videos if they watched at least one of the videos assigned to the week.



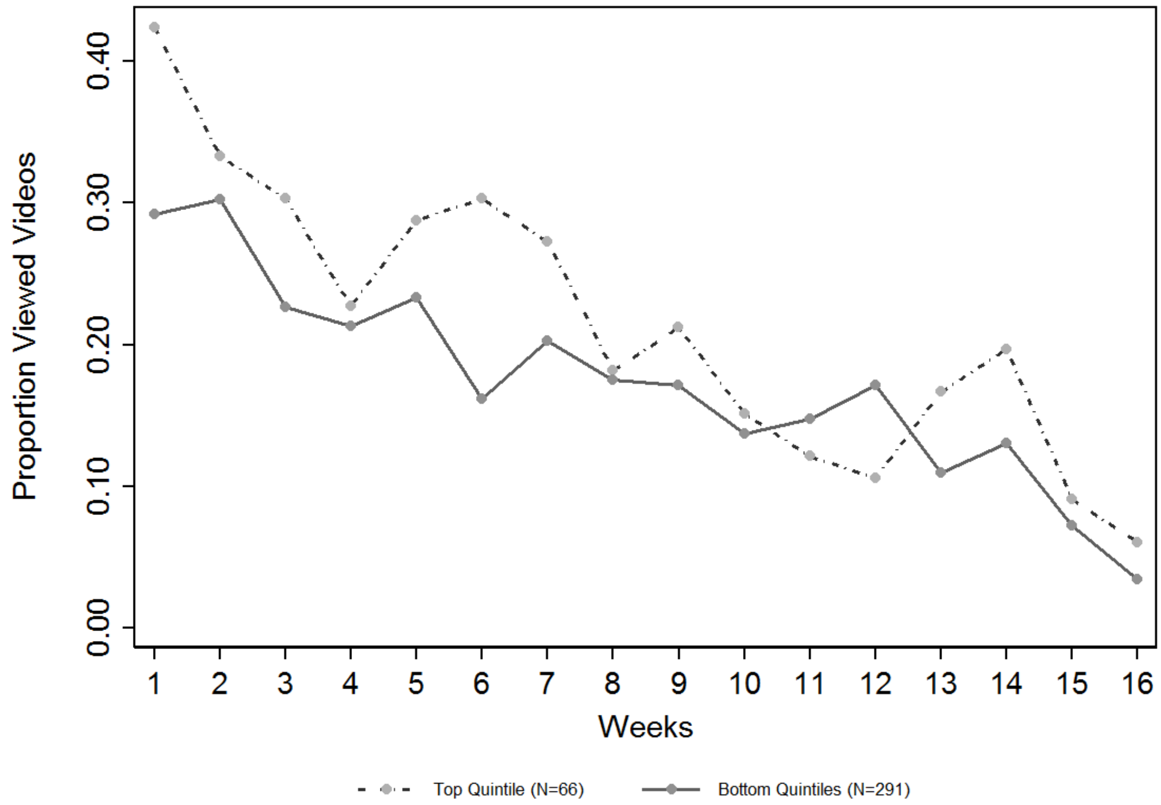
Appendix Figure 1. Video lecture screen shot

Notes: This is an example screen shot of an online video lecture from the Engineering Mathematics course. The professor's face is blurred to preserve anonymity.



Appendix Figure 2. Distribution of final course score, by source

Notes: This figure shows the distribution of students' final course scores for the students for which instructors directly provided course scores (N=578), by the source of the score: the registrar and the instructor.



Appendix Figure 3. Proportion of students viewing videos, by pretreatment GPA quintiles

Notes: This figure shows the proportion of students viewing the videos associated with each week's content by pretreatment GPA quintiles.

Table 1. List of participating courses by semester and professor

Course-Semester-Professor	Course	Semester	Professor	Recruited	Declined	Consented	Treatment	Control
1	Basics of Web Design	1	Professor 1	46	3	43	22	21
2	Computer Networking	1	Professor 2	86	2	84	43	41
3	Computer Organization	1	Professor 3	34	1	33	17	16
4		2	Professor 4	27	6	21	11	10
5	Computer Programming	1	Professor 5	23	0	23	13	10
6	Electronic Devices	1	Professor 6	26	4	22	12	10
7	Electronics Fundamentals	1	Professor 7	68	11	57	29	28
8		1	Professor 8	73	16	57	29	28
9		2	Professor 7	77	11	66	32	34
10		2	Professor 8	113	19	94	47	47
11	Engineering Mathematics	1	Professor 9	54	4	50	25	25
12		1	Professor 10	36	4	32	17	15
13		2	Professor 9	82	33	49	25	24
14		2	Professor 10	82	13	69	35	34
Total				**827**	**127**	**700**	**357**	**343**

Table 2. Student pretreatment characteristics, by treatment assignment

	Control			Treatment			diff	diff	
Characteristic	*mean*	*sd*	*n*	*mean*	*sd*	*n*	(T-C)	se	
Panel A: Student Characteristics									
Age at baseline	19.39	1.63	337	19.45	1.63	353	0.05	0.11	
Female	0.41	0.49	343	0.42	0.49	357	0.01	0.01	
Ethnic minority	0.16	0.37	337	0.16	0.37	353	0.00	0.03	
From Ulaanbaatar	0.39	0.49	329	0.36	0.48	346	-0.03	0.04	
Works for pay	0.13	0.33	329	0.10	0.29	346	-0.03	0.02	
Panel B: Educational Characteristics									
First year in program	0.12	0.33	341	0.12	0.32	355	-0.01	0.01	
Number of years enrolled at university	1.99	1.20	340	1.91	1.05	350	-0.09	0.06	
Engineering school	0.86	0.35	340	0.86	0.35	350	0.00	0.02	
Secondary school GPA of A	0.70	0.46	317	0.68	0.47	333	-0.03	0.04	
Total pretreatment credits	48.39	33.89	315	46.45	28.38	311	-2.71	1.88	
Pretreatment GPA	28.03	7.85	315	28.45	6.73	311	0.48	0.44	
Panel C: Household Characteristics									
Mother has bachelor's or higher	0.48	0.50	294	0.49	0.50	322	0.01	0.04	
Father has bachelor's or higher	0.34	0.48	268	0.37	0.48	287	0.03	0.04	
Household monthly income less than $425 USD	0.56	0.50	249	0.55	0.50	259	-0.01	0.05	
Household owns home	0.93	0.26	329	0.91	0.28	346	-0.01	0.02	
Household owns automobile	0.67	0.47	329	0.71	0.46	346	0.05	0.04	
Household owns refrigerator	0.92	0.28	329	0.94	0.23	346	0.03	0.02	
Household owns TV	0.94	0.24	329	0.96	0.20	346	0.02	0.02	
Panel D: Technology Access / Experience									
Access to computer at home	0.88	0.33	329	0.94	0.24	346	0.06	0.02	***
Access to internet at home	0.88	0.32	329	0.92	0.27	346	0.04	0.02	*
Has mobile phone with internet access	0.83	0.38	333	0.84	0.37	350	0.02	0.03	
Number of hours on computer in last 48 hours	9.60	7.83	329	9.73	7.71	346	0.07	0.59	
Taken course using lecture videos previously	0.77	0.42	329	0.79	0.41	346	0.02	0.03	
Previously enrolled in online course	0.67	0.47	329	0.68	0.47	346	0.01	0.04	
Previously completed online course	0.22	0.41	329	0.19	0.39	346	-0.04	0.03	
Panel E: Course Characteristics									
Course required for degree	0.88	0.33	337	0.90	0.30	353	0.02	0.02	
Somewhat or very interested in course	0.44	0.50	336	0.41	0.49	353	-0.04	0.04	
Somewhat or very familiar with course content	0.49	0.50	336	0.49	0.50	353	0.00	0.03	
Taken course previously	0.30	0.46	336	0.25	0.44	353	-0.05	0.03	*
Joint test (p-value) - All variables	0.61								
Joint test (p-value) - Panel A variables	0.22								
Joint test (p-value) - Panel B variables	0.89								
Joint test (p-value) - Panel C variables	0.65								
Joint test (p-value) - Panel D variables	0.47								
Joint test (p-value) - Panel E variables	0.35								

Notes: Table shows the means and standard deviations of student baseline characteristics. For binary characteristics, the proportion of students with the characteristic is shown. The treatment-control difference is the coefficient from a regression of the dependent variable on an indicator variable for treatment and randomization strata (i.e., course by wave by professor) fixed effects. Thus, the difference shown is not exactly equal to the difference between the treatment and control means shown. Results are robust to omitting the strata fixed effects. Standard errors are clustered at the school level. Robust standard errors shown. ***p<0.01, ** p<0.05, * p<0.1

Table 3. Impact on withdrawal and failing rates

	Score of zero				Score < 60			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.053*	0.048	0.059**	0.056*	-0.003	-0.005	0.011	0.011
	(0.031)	(0.029)	(0.028)	(0.029)	(0.039)	(0.037)	(0.035)	(0.036)
Control mean	0.165				0.460			
Control sd	0.371				0.499			
Observations	657	657	657	657	657	657	657	657
R-squared	0.005	0.159	0.239	0.282	0.000	0.149	0.249	0.280
Strata FE		yes	yes	yes		yes	yes	yes
Pretreatment GPA			yes	yes			yes	yes
Pretreatment covariates				yes				yes

Notes: This table shows linear probability models estimated using OLS. The outcome variables are a final course score (on a 100 point scale) of zero and a score less than 60. Robust standard errors shown. Missing valued of pretreatment GPA and additional covariates imputed using mean of nonmissing observations. Additional pretreatment covariates include those shown in Table 2. ***p<0.01, ** p<0.05, * p<0.1

Table 4. Impact on raw and standardized course score

	Course score (including zeros)				Course score (winsorized)			
					(p25)	(p50)	(p75)	(p90)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Raw Score								
Treatment	-1.680	-1.413	-2.620	-2.707	-2.184	-1.210	-0.225	-0.110
	(2.526)	(2.534)	(2.356)	(2.427)	(2.106)	(0.836)	(0.341)	(0.132)
Control mean	50.034				52.439	69.735	82.078	90.472
Control sd	34.310				30.950	12.321	4.734	1.840
Observations	657	657	657	657	657	657	657	657
R-squared	0.135	0.173	0.292	0.330	0.294	0.298	0.183	0.116
Panel B: Standardized Score								
Treatment	-0.034	-0.025	-0.061	-0.062	-0.085	-0.043	-0.022	-0.011
	(0.078)	(0.079)	(0.073)	(0.076)	(0.063)	(0.028)	(0.013)	(0.007)
Control mean	0.017				0.116	0.582	0.940	1.231
Control sd	1.022				0.864	0.393	0.187	0.094
Observations	657	657	657	657	657	657	657	657
R-squared	0.000	0.045	0.179	0.222	0.178	0.174	0.131	0.099
Strata FE		yes	yes	yes	yes	yes	yes	yes
Pretreatment GPA			yes	yes	yes	yes	yes	yes
Pretreatment covariates				yes				

Notes: Panel A shows raw final course scores obtained from registrar office (on scale of 0 to 100). Panel B shows final course scores obtained from registrar office, standardized within course-professor-semester grouping. Winsorized regressions replace zeros and scores at or below the 25th, 50th, 75th and 90th percentile score in the control distribution with the control group score at the respective percentile of interest. Robust standard errors shown. Pretreatment covariates include those shown in Table 2. Missing values imputed using mean of nonmissing across covariates. ***p<0.01, ** p<0.05, * p<0.1

Table 5. Treatment heterogeneity, by pretreatment GPA

	Score of zero	Score < 60	Std. Score
	(1)	(2)	(3)
Treatment	0.062*	-0.029	-0.016
	(0.036)	(0.042)	(0.087)
Treatment * top quintile	-0.079*	0.084	0.019
	(0.048)	(0.088)	(0.186)
Top quintile	-0.114**	-0.360***	0.702***
	(0.049)	(0.074)	(0.157)
Control mean (lower quintiles)	0.200	0.528	-0.072
Control sd (lower quintiles)	0.401	0.500	1.024
Observations	657	657	657
R-squared	0.174	0.186	0.089

Notes: Linear probability model estimated using OLS. Final course score (scale of 0 to 100) obtained from registrar used. Robust standard errors shown. All models include strata fixed effects. Missing pretreatment GPA imputed using mean of nonmissing observations. Results are robust to disaggregating students by pretreatment terciles and quartiles, nonimputation and exclusion of covariates. ***p<0.01, ** p<0.05, * p<0.1

Table 6. Post-treatment course completion, persistence, and post-treatment GPA

	Left Program	Post Credits	Post GPA	Passed Focal	Retook Focal	Retake Pass
	(1)	(2)	(3)	(4)	(5)	(7)
Treatment	0.013	1.235	-0.256	-0.002	0.054*	0.037
	(0.022)	(0.974)	(0.720)	(0.033)	(0.031)	(0.025)
Control mean	0.099	33.848	24.080	0.708	0.193	0.109
Control sd	0.300	13.886	9.885	0.455	0.395	0.312
Observations	657	529	529	657	657	657
R-squared	0.196	0.354	0.306	0.219	0.106	0.102

Notes: Linear probability model estimated using OLS. Post-treatment outcomes based on transcript data in the two years post intervention. "Passed focal course" indicates that a student passed the focal class of the study (by score 60 or higher) at some point in the semesters 2 years post intervention. "Dropped out" indicates the student left the program without a degree in the 2 years post treatment - all students with a registrar status including the following: dropped out, expelled, inactive, status unknown. Post-treatment GPA calculated by dividing course scores (on 100 point scale) by total units enrolled in post treatment (range: 0 to 70). All models include strata fixed effects and pretreatment GPA. Missing pretreatment GPA imputed using mean of nonmissing observations. Results are robust to nonimputation and exclusion of covariates. ***p<0.01, ** p<0.05, * p<0.1

Table 7. Compliance and course experience

	treatment	*se*	*p*		control mean	control sd
Panel A: Compliance (Platform, N=657)						
Logged onto platform and watched at least 1 video	0.60	0.03	0.00	***	0.00	0.00
Number of videos viewed	7.39	0.71	0.00	***	0.00	0.00
Number of weeks' videos viewed	2.95	0.20	0.00	***	0.00	0.00
Panel B: Compliance (Self-Reported, N=551)						
Reported receiving access to videos	0.79	0.03	0.00	***	0.08	0.27
Able to access course videos	0.67	0.03	0.00	***	0.23	0.42
Reported watching official course videos in a typical week	0.58	0.03	0.00	***	0.15	0.36
Minutes watching official course videos in a typical week	107.71	14.44	0.00	***	11.18	30.20
Panel C: Course Activity in Typical Week (Self-Reported, N=551)						
Reported attending in-person section in a typical week	-0.27	0.03	0.00	***	0.94	0.25
Number of weeks of in-person section attended	-1.51	0.34	0.00	***	13.13	3.66
Minutes attending in-person section in a typical week	-52.59	21.18	0.01	**	167.86	247.37
Reported meeting with professor 1-on-1	0.02	0.04	0.58		0.29	0.46
Minutes meeeting with professor 1-on-1	8.34	8.85	0.35		22.53	66.72
Reported studying alone	-0.04	0.04	0.39		0.59	0.49
Minutes studying alone	0.02	13.53	1.00		85.38	180.88
Reported studying with peers	-0.05	0.04	0.19		0.32	0.47
Minutes studying with peers	-7.47	9.53	0.43		43.30	119.71
Reported completing assignments alone	-0.05	0.03	0.16		0.82	0.39
Minutes completing assignments alone	4.91	17.79	0.78		133.70	184.39
Reported completing assignments with peers	-0.06	0.04	0.12		0.42	0.50
Minutes completing assignments with peers	-19.01	12.83	0.14		64.03	199.25
Reported watching other online tutorias	0.01	0.04	0.84		0.37	0.48
Minutes watching other online tutorial	30.77	15.36	0.05	**	34.25	78.85
Panel D: Course Satisfaction (Self-Reported, N=551)						
Found in-person section useful	-0.28	0.03	0.00	***	0.89	0.31
Finds in-person interaction with professor very important	-0.02	0.04	0.71		0.59	0.49
More interested in topic after course	-0.06	0.03	0.07	*	0.87	0.33
More likely to take next course in sequence	0.03	0.04	0.38		0.74	0.44
Interested in taking a future course with lecture videos	0.02	0.04	0.67		0.63	0.48
Satisfied (very or somewhat) in course exerpience	-0.02	0.04	0.54		0.80	0.40
Peers engaged (very or somewhat) in course experience	0.02	0.03	0.57		0.81	0.39

Notes: Outcomes from platform data (N=657) and endline student survey (n=551). Models estimated using OLS, controlling for strata fixed effects, and using robust standard errors. ***p<0.01, ** p<0.05, * p<0.1

Table 8. Classroom observation comparisons

	Control (N=16)		Treatment (N=13)		diff	diff	
	mean	*sd*	*mean*	*sd*	(T-C)	se	
Panel A: Observation characteristics							
Session length (minutes)	89.38	7.80	57.62	21.15	-31.76	6.16	***
Original enrollment class size	36.31	18.34	29.23	11.35	-7.08	5.57	
Number of students in attendance	20.44	9.67	9.92	5.12	-10.51	2.81	***
Proportion of students in attendance	0.59	0.16	0.35	0.16	-0.24	0.06	***
Professor is primary instructor	1.00	0.00	0.77	0.44	-0.23	0.12	*
More than half of students unengaged	0.81	0.40	0.39	0.51	-0.43	0.17	**
Instructor unorganized	0.06	0.25	0.23	0.44	0.17	0.14	
Students mentioned online videos	0.13	0.34	0.92	0.28	0.80	0.12	***
Instructor mentioned online videos	0.13	0.34	0.92	0.28	0.80	0.12	***
Panel B: Proportion of time on classroom activities							
Attendance	0.02	0.04	0.09	0.20	0.07	0.06	
Classroom management	0.02	0.04	0.06	0.09	0.04	0.03	
Teaching new material	0.63	0.29	0.00	0.00	-0.63	0.07	***
Reviewing online video	0.00	0.00	0.22	0.33	0.22	0.09	**
Reviewing prior week material	0.04	0.10	0.26	0.41	0.22	0.12	*
Instructor led Q&A	0.06	0.09	0.24	0.29	0.18	0.08	**
Students working independently	0.02	0.07	0.03	0.09	0.01	0.03	
Students working in groups	0.00	0.00	0.00	0.00	0.00	0.00	
Students completing quiz	0.06	0.08	0.10	0.24	0.04	0.07	
Students completing exam	0.00	0.00	0.00	0.00	0.00	0.00	
Students giving presentation	0.03	0.10	0.01	0.04	-0.01	0.03	
Instructor present but off task	0.00	0.01	0.00	0.00	0.00	0.00	
No instructor in classroom	0.04	0.06	0.04	0.06	0.00	0.02	
Other	0.05	0.08	0.09	0.21	0.04	0.06	

Notes: Table shows the means and standard deviations of characteristics recorded during course observations and the proportion of time recorded spent on various classroom activities. In total, 29 observations were recorded (one observation all treatment and control sections for each course-professor-semester offering, with the exception of treatment observations for two courses for which instructors did not regularly hold in-person sessions). Differences shown are a simple difference with robust standard errors shown. ***p<0.01, ** p<0.05, * p<0.1

Table A1. Attrition rate, by treatment assignment

	Control			Treatment			diff	diff	
Attrition Type	*mean*	*sd*	*n*	*mean*	*sd*	*n*	(T-C)	se	
Missing registrar score or endline survey	0.24	0.43	343	0.19	0.39	357	-0.04	0.03	
Missing registrar score	0.06	0.24	343	0.06	0.24	357	0.00	0.02	
Missing endline survey (not found or declined)	0.22	0.41	343	0.16	0.37	357	-0.05	0.03	*
Not found for endline survey	0.16	0.36	343	0.11	0.31	357	-0.04	0.03	*
Found but declined endline survey	0.08	0.27	290	0.06	0.24	318	-0.01	0.02	

Notes: Notes: This table shows mean attrition for missing registrar and endline survey outcomes. The treatment-control difference reported is the coefficient from a regression of the dependent variable on an indicator variable for treatment and randomization strata (i.e., course by wave by professor) fixed effects. Thus, the difference shown is not exactly equal to the difference between treatment and control means shown. Results are robust to omitting the strata fixed effects. Robust standard errors shown. ***p<0.01, ** p<0.05, * p<0.1

Table A2. Compliance and course experience among students receiving a zero

	treatment	*se*		control mean	control sd
Panel A: Compliance (Platform, N=126)					
Logged onto platform and watched at least 1 video	0.37	0.07	***	0.00	0.00
Number of videos viewed	1.55	0.45	***	0.00	0.00
Number of weeks' videos viewed	0.86	0.24	***	0.00	0.00
Reported receiving access to videos	0.78	0.09	***	0.07	0.25
Able to access course videos	0.78	0.09	***	0.07	0.25
Panel B: Compliance (Self-Reported, N=75)					
Reported watching official course videos	0.69	0.10	***	0.07	0.25
Minutes watching official course videos	69.21	11.17	***	5.67	21.61
Reported attending in-person section	-0.61	0.13	***	0.87	0.35
Number of weeks of in-person section attended	-1.92	1.37		7.50	4.39
Minutes attending in-person section	-79.02	34.56	**	105.83	139.61
Panel C: Course Activity in Typical Week (Self-Reported, N=75)					
Reported meeting with professor 1-on-1	0.00	0.00		0.03	0.18
Minutes meeeting with professor 1-on-1	0.00	0.00		0.67	3.65
Reported studying alone	-0.41	0.15	***	0.47	0.51
Minutes studying alone	-16.96	41.65		55.10	76.43
Reported studying with peers	-0.03	0.04		0.07	0.25
Minutes studying with peers	-3.97	4.93		6.00	24.16
Reported completing assignments alone	-0.12	0.15		0.70	0.47
Minutes completing assignments alone	-19.29	23.40		79.17	64.84
Reported completing assignments with peers	-0.22	0.12	*	0.27	0.45
Minutes completing assignments with peers	-21.03	14.76		41.00	86.48
Reported watching other online tutorias	-0.11	0.15		0.17	0.38
Minutes watching other online tutorial	0.79	16.91		18.00	44.98
Panel D: Course Satisfaction (Self-Reported, N=75)					
Found in-person section useful	-0.77	0.09	***	0.93	0.25
Finds in-person interaction with professor very important	-0.22	0.17		0.47	0.51
More interested in topic after course	0.00	0.14		0.70	0.47
More likely to take next course in sequence	0.03	0.14		0.80	0.41
Interested in taking a future course with lecture videos	0.13	0.16		0.67	0.48
Satisfied (very or somewhat) in course exerpience	-0.02	0.14		0.40	0.50
Peers engaged (very or somewhat) in course experience	0.07	0.09		0.87	0.35

Notes: Outcomes from platform data (N=126) and endline student survey (n=75) for just students receiving a course score of zero. Models estimated using OLS, controlling for strata fixed effects, and using robust standard errors. ***p<0.01, ** p<0.05, * p<0.1

Table A3. Impact on scores at threshold grades

	Score greater than or equal to				
	60	70	80	90	100
Treatment	-0.011	-0.067**	-0.023	-0.007	-0.017*
	(0.035)	(0.033)	(0.031)	(0.023)	(0.009)
Control mean	0.540	0.410	0.273	0.102	0.022
Control sd	0.499	0.493	0.446	0.304	0.146
Observations	657	657	657	657	657
R-squared	0.249	0.308	0.249	0.116	0.174
Strata FE	yes	yes	yes	yes	yes
Pretreatment GPA	yes	yes	yes	yes	yes
Pretreatment covariates					

Notes: This table shows linear probability models estimated using OLS. The outcome variables are receiving a final course score (on a 100 point scale) greater than or equal to the threshold scores of 60, 70, 80, 90, and 100. Robust standard errors shown. Missing valued of pretreatment GPA imputed using mean of nonmissing observations. Results robust to exclusion of pretreatment GPA and inclusion of additional pretreatment covariates shown in Table 2. ***p<0.01, ** p<0.05, * p<0.1