Disconnecting Women:
Gender Disparities in the Impact of Online Instruction*

Xiaoyue Shan
(National University of Singapore)

Ulf Zölitz
(University of Zurich)

Uschi Backes-Gellner
(University of Zurich)

May 20, 2025

Abstract

We study the impact of online instruction with a field experiment that randomly assigns 1,344 university students to different proportions of online and in-person lectures in multiple introductory courses. Increased online instruction leaves men's exam performance unaffected but significantly lowers women's performance, particularly in math-intensive courses. Online instruction also reduces women's longer-run performance and increases their study dropout. Exploring mechanisms, we find that women exposed to more online lectures report greater difficulty in connecting with peers, less engaging instructors and lower course satisfaction. Our findings caution policymakers that shifting toward more online instruction may disproportionally harm women.

Keywords: Online instruction, field experiment, gender disparities

JEL Codes: J16, I23, C93

* We received helpful comments from Jessica Pan, Jan Feld and participants at the Family and Education Workshop (FEW), the Advances with Field Experiments (AFE) conference, the National University of Singapore, the University of Zurich, Peking University, and Singapore Management University. We gratefully acknowledge financial support from the University of Zurich URPP Equality of Opportunity. We thank Matthew Bonci, Raffaella Dimastrochicco, Madeleine Smith and Albert Thieme for excellent research support.

1 Introduction

The rapid expansion of online instruction, accelerated by the Covid-19 pandemic, has reshaped the landscape of higher education. By 2024, over 95% of post-secondary institutions in the U.S. have offered some form of online instruction. Within these institutions, 40% of college administrators reported plans to increase spending on online education, compared to only 2% intending to reduce it (Venable et al., 2024). While online education appeals for its flexibility and cost-effectiveness (Deming et al., 2015; Goodman et al., 2022), it substantially reduces the social and interactive elements of education. Introducing online education also imposes greater challenges for students and families less engaged with online technologies. A growing body of literature highlights that "zooming to class" negatively impacts student performance (e.g., Figlio, Rush and Yin, 2013; Kofoed et al., 2024). This paper presents novel evidence that online instruction disproportionately harms women's short- and long-term educational attainment.

Why would online instruction affect men and women differently? Prior studies suggest that women are more likely than men to be social learners and rely more on personal interactions with faculty and peers throughout the learning process (Arbaugh, 2000; Kim and Sax, 2009; Wong and Chapman, 2023). Ample evidence also shows that men and women differ in their access, use, and engagement with digital technology—a phenomenon called the "gender digital divide."[1] The expansion of online education may therefore disadvantage women by reducing social interactions and increasing the use of information technologies and online tools.

In this paper, we examine how online instruction differentially affects the learning outcomes of men and women. We conducted a field experiment among 1,344 students enrolled in multiple introductory courses at an international business school during the pandemic. To offer in-person learning opportunities while adhering to social distancing regulations, we introduced a "partial attendance policy" in collaboration with the business school. This policy created experimental variation in the teaching mode by randomly assigning students to different proportions of online versus in-person lectures for each course. We exploit this within-student, cross-course variation to analyze how online instruction affects students' performance, learning

[1] Globally, 58% of men versus 48% of women report internet usage (ITU, 2019). Women are also substantially less likely than men to use AI tools like ChatGPT (Humlum and Vestergaard, 2024; Carvajal, Franco and Isaksson, 2024). Men also tend to be more confident about their digital skills (Jackson et al., 2010; Li and Kirkup, 2007) and use the internet for solitary purposes, whereas women more often use it for socializing and communication (Weiser, 2000; Joiner, Stewart and Beaney, 2015; Tsai and Tsai, 2010).

experiences, and study persistence. The experiment was implemented in the first three weeks of the semester, a crucial onboarding period for first-year students. Although the experimental period is brief—the policy was terminated due to the worsening of the pandemic situation—we find lasting impacts of the instruction mode on student outcomes.

Our results show that being assigned to more online lectures has an overall negative effect on students' exam performance, consistent with most prior work. However, this overall effect masks large gender differences: We find that increasing the proportion of online lectures by 10 percentage points lowers women's exam score by 0.04 SD (p-value = 0.0002). For men, in contrast, we observe a precisely estimated null effect. The 95% confidence intervals allow us to rule out effects smaller than −0.01 SD and larger than 0.015 SD when men's online share increases by 10 percentage points.

Could this gender gap in the impact of online instruction be a chance finding? Given that gender is a frequently observed and commonly tested characteristic, we must worry about reporting a statistically significant yet spurious gender gap. This concern is especially valid for the first paper documenting a stark gender difference in the effect of online instruction. To rule out the possibility of a false positive, we apply multiple testing corrections to all pre-registered dimensions of heterogeneity. Applying the step-down procedure by Romano and Wolf (2005, 2016) yields a corrected p-value of 0.001 (original p-value = 0.0002). A more conservative Bonferroni correction yields a p-value of 0.004. These results clearly reject the idea of a chance finding.

After establishing the negative impact of online instruction on women's overall exam performance, we unpack this impact by linking lecture content to individual exam questions. We then test how online instruction affects student performance for lecture content covered during and after the experimental period. We find that online instruction does not affect women's performance on exam questions related to content taught early in the term—basic yet fundamental concepts. However, online instruction significantly reduces their performance for lecture content covered later in the semester—more advanced and difficult topics building upon the earlier content. This negative dynamic spillover effect may be due to (1) women not fully grasping the basic content taught online or (2) women's interest and later investments decreasing after the initial online exposure. This dynamic effect suggests that online instruction may be particularly harmful for women's longer-term learning outcomes.

Next, we investigate whether the impact of online lectures varies across courses. Our results show that the negative effect of online instruction on women is strongest in the two most math-intensive courses: Mathematics and Microeconomics. Specifically, a 10-percentage-points increase in the proportion of online lectures in these courses lowers women's final course grades by approximately 0.08 SD—compared to only 0.02 SD in other courses. One possible explanation for this result could be that the lack of immediate feedback, clarification opportunities, and peer support is particularly harmful for math-intensive subjects where women are less self-confident than men (Herbert & Stipek, 2005; Else-Quest et al., 2010).

While our results clearly show that online instruction adversely affects women's short-term performance, it is unclear whether these effects translate into any meaningful long-term disadvantages. To be able to estimate longer-term effects on student-level outcomes such as study dropout, we use student-level variation in online instruction in the two math-intensive courses.[2] The results show that when exposed to more online lectures in these courses, women earn fewer credits and become more likely to drop out of their studies in the following years. A 10-percentage-points increase in online lectures raises women's study dropout probability by 0.07 percentage points or 19%. This substantial effect highlights that exposure to online instruction can have lasting negative effects on women's educational careers.

Next, we examine why online instruction negatively affects women's performance but has no impact on men. We use survey data and attendance records of online lectures to explore two broad mechanisms: (1) gender differences in the preference for online lectures; and (2) how online instruction affects learning experiences.

To understand preferences for online instruction, we survey students about their preferred teaching mode at the end of the semester. We find that women are 36% less likely than men to report online instruction as their preferred mode. This gender difference is also reflected in lecture attendance behavior: Overall, women are less likely than men to attend online lectures, and this gender gap is substantially larger for lectures for which in-person attendance is permitted. This

[2] Students' longer-term outcomes do not vary at the course level, so we must define a proxy for online exposure that varies at the student level. One way to define student-level variation is to calculate the average proportion of online lectures in all courses. However, by design, our partial attendance policy does not create exogenous variation in the overall exposure to online lectures: We wanted to give all students equal opportunities to access in-person lectures—although, unavoidably, the exposure was not equal across courses. Given our finding that the effect of online instruction is concentrated in the two math-intensive courses (Mathematics and Microeconomics), we focus on how online lectures in these courses affect students' long-run outcomes.

suggests that, whenever physical attendance is possible, women are more likely to attend the lectures in person (instead of attending remotely)—confirming their stronger preference for in-person instruction.[3]

To understand how online instruction affects students' learning experiences, we use endline survey questions that ask students to evaluate different course aspects. Our results show that women exposed to more online lectures report (1) lower satisfaction with the course and its lectures, (2) increased difficulties in making peer-to-peer contact, and (3) the professor being less engaging and responsive. Men, in contrast, appear either unaffected or, for some course aspects, even more satisfied when assigned to more online lectures. These results suggest that online instruction disproportionally harms women's performance because it prevents them from effectively engaging with instructors and peers.

In the final part of our paper, we conduct a meta-analysis of the literature that examines the impact of online instruction in higher education.[4] The aim of this meta-analysis is to situate our estimates into the broader literature and to explore why previous work has overlooked gender differences in the response to online learning. Our results show that there is a broad consensus on the overall negative impact of online instruction with an average effect of -0.20 SD (p-value = 0.0006). However, our main finding—that women experience disproportionately negative impacts from online education—has not been documented before. The most plausible reason why this finding has remained undocumented is statistical power. With a median sample size of 491 observations, most prior studies simply lack the power to conclusively test gender differences.

Our study makes two main contributions to the literature. (1) We provide the first experimental evidence that online instruction disproportionately harms women. (2) We deliver comprehensive results on how online instruction affects short- and longer-term educational attainment, attendance behaviors, and student engagement. While previous studies have looked at the contemporaneous effects of online education, we provide novel evidence on longer-term effects and mechanisms. Our results provide a warning for educators and policymakers. While online education offers greater accessibility, flexibility, and cost-saving potentials in times of

[3] When examining students' overall course attendance, we find that women assigned to more online lectures report lower attendance rates. This suggests that online instruction may reduce women's exam performance by discouraging them from attending lectures.
[4] Our meta-analysis includes 10 prior studies: Bettinger et al., 2017; Coates et al., 2004; Xu and Jaggars, 2013; Figlio, Rush and Yin, 2013; Bowen et al., 2014; Joyce et al., 2015; Alpert, Couch and Harmon, 2016; Cacault et al., 2021; Alegría et al., 2023; Kofoed et al., 2024.

shrinking university budgets, it appears to do so at the expense of reducing women's performance, engagement, satisfaction, and ultimately degree completion rates.

Our study joins a growing body of research underscoring the importance of social interaction—and the potential downsides of digital technologies—for women in both educational and workplace settings. Prior work shows that women benefit more than men from interactive, collaborative learning environments (Belenky et al., 1988; Klein et al., 1994; Arbaugh, 2000; Zhu, 2012), are less likely to adopt AI tools (Carvajal et al. 2024; Humlum and Vestergaard, 2024), earn less on online labor platforms (Litman et al., 2020; Cook et al., 2021; Adams-Prassl et al., 2023), and gain more from close physical proximity to teammates (Emanuel, Harrington and Pallais, 2023). We extend this literature by demonstrating that women show a stronger preference for face-to-face instruction and experience larger performance gains from in-person, interactive teaching than men.

2 The Experiment

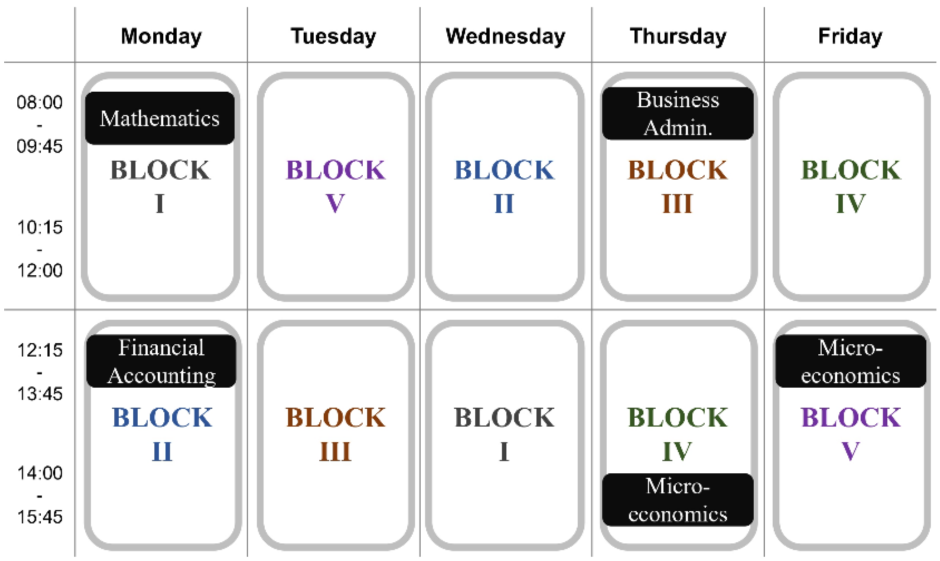
Institutional context: Our experiment was conducted at an international business school that is part of a large public European research university. Each year, the business school enrolls over 1,000 students who pursue major or minor programs in business, economics, finance, or informatics. To major in these fields, students must successfully complete a series of mandatory first-year courses. Our experiment involves four such compulsory courses, typically taken in the first semester: Microeconomics (9 credits), Mathematics (6 credits), Financial Accounting (6 credits), and Business Administration (3 credits).

Pandemic background: In the spring semester of the 2019/20 academic year, the Covid-19 pandemic broke out. As a result, the university abruptly switched to online instruction. After observing declining infection numbers during the summer of 2020, the university decided to reintroduce some in-person instruction in fall 2020, subject to social distancing regulations. In collaboration with the business school, we introduced a "partial attendance" policy that allowed students to attend a random subset of lectures in person.

Partial attendance experiment: All incoming first-year bachelor's students at the business school were randomly assigned to one of five attendance groups (Groups A–E). Each group was allocated to one of five distinct teaching blocks, limiting each student's physical lecture attendance to their assigned block. This setup limited campus attendance to 20% of students at any given time, thereby adhering to social distancing regulations. Figure 1(a) illustrates the weekly timetable, the division into five teaching blocks, and the scheduled lectures within each course.

Figure 1. The Weekly Timetable and the Schedule for In-Person Attendance

(a) Teaching Blocks and Courses



(b) In-Person Attendance Schedule

Attendance Group (A–E)

Teaching Block (I–V)	A	B	C	D	E	
	I	II	III	IV	V	(Week 1)
	II	III	IV	V	I	(Week 2)
	III	IV	V	I	II	(Week 3)

Notes: Figure (a) shows the timetable of an instruction week and the scheduled times of four courses in our sample: Mathematics, Financial Accounting, Business Administration, and Microeconomics. Figure (b) shows which teaching blocks students in each attendance group were allocated to from week 1 to week 3. When assigned to a teaching block, students were permitted to physically attend courses in the block. For example, students in Group A can physically attend courses in Block I in week 1, courses in Block II in week 2, and courses in Block III in week 3.

Attendance blocks rotated on a weekly level, creating variation in the availability of in-person lectures at the course level. Figure 1(b) provides an overview of how attendance groups rotated through the teaching blocks during the first three weeks of the semester. For instance, students in Group A can physically attend lectures in Block I (Mathematics) during Week 1, Block II (Financial Accounting) in Week 2, and Block III (Business Administration) in Week 3. Similarly, students in Group B attended Financial Accounting in Week 1, Business Administration in Week 2, and Microeconomics in Week 3. Groups C, D, and E rotated in the same way.[5]

Although the partial attendance policy was designed to cover the whole fall semester, it only lasted for three weeks due to unexpected changes in university policies and the Covid-19 situation. From week 4, the university relaxed restrictions on in-person instruction given the low infection rate during the past weeks. However, the Covid situation rapidly deteriorated in October, and the university abandoned all in-person instruction from week 7 onwards. All remaining teaching and exams were administered online. Appendix Figure A.1 shows a timeline of our experiment and the Covid situation.

Variation in online instruction: The partial attendance policy introduced experimental variation in the share of online lectures during the first three weeks of the semester which we will exploit for identification. The policy introduced variation in online lectures at the student-by-course level. For instance, students in Group A could physically attend one out of three lectures in the Mathematics, Financial Accounting, and Business Administration courses, but had to attend all Microeconomics lectures online. In contrast, students in Group D were permitted to attend two out of six Microeconomics lectures in person but had to attend all Financial Accounting and Business Administration lectures remotely. Among all 3,770 student-by-course observations in our sample, the proportion of online lectures ranges from 66.7% to 100%, with an average value of 80%.

Pre-registration: Our experiment received IRB approval from the University of Zurich and was pre-registered on the AEA RCT Registry (ID: AEARCTR-0006538). Our pre-registration

[5] The timetable also includes other courses, depending on students' study program and year of study. Almost all courses offered at the business school had to follow the assigned rotating schedule: in-person attendance was only permitted for certain students in designated weeks. However, many instructors decided to deliver their course purely or primarily online. Therefore, we focus our analysis on the four large compulsory courses that strictly followed the partial attendance schedule.

specifies four research questions on how online instruction affects (1) question-level performance, (2) course-level performance, (3) long-term outcomes, and (4) learning experiences. While we present findings on all four domains, we place less emphasis on the question-level analysis than originally planned—due to early termination of the experiment. Additionally, although we preregistered gender as one of the nine dimensions of effect heterogeneity, we did not anticipate finding substantial gender differences. Given the strong gender disparities uncovered, this paper primarily explores these gender differences, while still examining heterogeneity across all other pre-specified student attributes. We apply multiple testing corrections to these heterogeneity analyses, as we will detail in Section 5.2. In Appendix B, we discuss these deviations from the pre-analysis plan in greater detail.

3 Data and Descriptive Statistics

We use administrative university data, Zoom attendance data, as well as baseline and endline survey data, to estimate the impact of online instruction on students' hard- and soft-outcomes. In the following, we briefly describe the different data sets.

Administrative data: We collected administrative data on students' gender, age, nationality, major, and home address. We also collected data on all original exam questions (mostly multiple-choice questions), students' obtained points for each individual question, their total exam points, as well as their final grade for the course. In the analysis, we primarily use the total exam score as the measure of course performance, since it is the most nuanced performance measure. To test robustness of our results, we also use the final grade and course passing indicator as alternative outcomes. To unpack the overall impact of online instruction on course-level performance, we further use question-level data to analyze impact on performance for different parts of the exam (see Section 5.3).

Zoom attendance data: To explore students' lecture attendance behaviors, we use the Zoom lecture attendance lists from the Microeconomics course, the largest course taken by most students in our sample. We use Zoom data to compare students' online attendance rate depending on whether in-person attendance was allowed and not.[6] Although the data is only available for one course, it helps to unveil systematic differences in students' revealed preferences for online or in-person lectures (see Section 6 for details).

Baseline survey data: Before the start of the term, we conducted a baseline survey among students enrolled in the Microeconomics course. The survey measures students' high school grades, personality traits, and study habits, and had a high response rate of 85%. Although the data are not available for all students in our sample, we use it to analyze effect heterogeneity in Section 5.2.

Endline survey data: At the end of the semester, we conducted an endline survey among all students subject to the partial attendance experiment. In this survey, we elicited students' preferences for future teaching modes and asked them to evaluate different aspects of each course they took. For example, students were asked to evaluate how much they liked a course, how much the instructor engaged with students and whether the students found it difficult to interact with their peers. The response rate for the endline survey is 53%, and Appendix Table A.9 shows that the assigned proportion of online lectures did not affect the response rate. We use the endline survey data to provide suggestive evidence on underlying mechanisms (see Section 6).

Descriptive statistics: Table 1 shows descriptive statistics for our student population and their exam performance. Panel A summarizes student characteristics observed in the administrative data. Among the 1,344 students in our sample, 41% are female; the average age is 21; 16% of them are international students (i.e., foreign nationals); and the median commuting time from home to the campus is 36 minutes. Regarding study program, 57% of students are enrolled in a business, economics, or finance major; 18.5% are enrolled in the informatics major; and the remaining 24.5% are enrolled in a minor program at the business school but major in a field at another faculty.

Panel B of Table 1 provides information about 892 students answering the baseline survey. The average high school final grade for math and language is 4.7 (on a scale of 1 to 6). Students also report their high school study hours studying alone and with others. We use the proportion of hours studying alone (77% on average) as a measure of self-study intensity. The survey also

[6] Our attendance data only covers the period of week 1 to week 7 of the semester because the course instructor missed the opportunity to download zoom attendance data for the final part of the semester.

measures students' general tendency for group study with the following questions: "Working together with other students helps me to understand the subject matter;" and "Learning on my own is more effective than working with others" (reversed). Students answer to what extent these statements apply to them on a Likert scale of 1 (does not apply at all) to 7 (applies perfectly), and we use the average response to these items as a proxy for students' group study preferences. Finally, we also elicit students' conscientiousness and extraversion using questions from the Short 15-item Big Five Inventory (Gerlitz and Schupp, 2005).

Table 1. Descriptive Statistics

	(1)	(2)	(3)	(4)	(5)
	N	Mean	SD	Min	Max
Panel A: Student Characteristics (from admin data)					
Male	1,344	0.592	0.492	0	1
Age	1,344	21.33	2.551	18	41
International Student	1,344	0.157	0.364	0	1
Commute Time to University (minutes)	1,344	40.55	37.08	4.25	600.55
Business/Economics/Finance Majors	1,344	0.570	0.495	0	1
Informatics Majors	1,344	0.185	0.388	0	1
Other Majors	1,344	0.246	0.431	0	1
Panel B: Student Characteristics (from baseline survey)					
Conscientiousness	892	4.796	0.955	1	7
Extraversion	892	4.747	1.244	1	7
High School Grade	892	4.663	0.512	2.75	6
High School Self-Study Intensity	880	0.770	0.201	0	1
Group Study Preference	890	3.949	1.250	1	7
Panel C: Performance					
Total Exam Score	3,770	34.199	18.887	0	106
Final Course Grade	3,770	4.348	0.838	1	6
Passing Course	3,770	0.753	0.431	0	1
Points for Individual Exam Questions	145,027	1.135	1.361	-1.5	4

Notes: The table shows summary statistics for students' background characteristics and their performance outcomes at the course level or exam question level. For each variable, we show the number of observations (*N*), the mean value, the standard deviation (SD), and the minimum and maximum values. Student characteristics observed for all 1,344 students are derived from administrative data; other variables are derived from a baseline survey with 892 respondents. *High School Grade* refers to the average final math and language grade in high school. *High School Self-Study Intensity* refers to the proportion of study hours spent alone (versus with other students). *Group Study Preference* is measured with two survey questions: "Working together with other students helps me to understand the subject matter" and "Learning on my own is more effective than working with others" (reversed)—both answered on a scale of 1 (does not apply at all) to 7 (applies perfectly). These two variables are missing for several students due to invalid responses. Appendix Table A.1 shows the summary statistics separately for men and women.

4 Estimation Strategy

We estimate the impact of online instruction on learning outcomes using the following model as our baseline specification:

$$Y_{ic} = \beta \cdot Online_{ic} + \alpha_c + \gamma \boldsymbol{X}_i + \varepsilon_{ic} \tag{1}$$

In this specification, the outcome variable Y_{ic} is the performance of student i in course c. Our analysis focuses on the total exam score as the primary performance outcome and uses final course grades and course passing indicators as alternative outcomes in robustness checks. For ease of interpretation, we standardize all non-binary outcome variables to have a mean of 0 and a SD of 1 across the estimation sample. The key treatment variable of interest, $Online_{ic}$, captures the exposure of student i to online instruction in course c (i.e., the proportion of lectures assigned to be online). We always control for course fixed effects, captured by α_c. We then augment the basic model by the vector $\boldsymbol{X}_i$ capturing student characteristics at baseline. $\boldsymbol{X}_i$ includes gender, age, an indicator for international students, log commute time from home to university, major category fixed effects, and an indicator for first-year students. ε_{ic} is the error term. Since we observe multiple courses per student, we cluster standard errors at the student level. Standard errors remain very similar when clustering at the student-by-course level.

One advantage of our empirical design is that the treatment varies at the student-by-course level, which means that we can include student fixed effect to absorb all student-level variation. Equation (2) shows this even more restrictive model, which represents our preferred specification:

$$Y_{ic} = \beta \cdot Online_{ic} + \alpha_c + \gamma_i + \varepsilon_{ic} \tag{2}$$

Here, γ_i represents student-level fixed effects. This model exploits within-student, cross-course variation in the proportion of online lectures to estimate the impact on online instruction, holding all observed and unobserved student characteristics constant.

To estimate the gender gap in the impact of online instruction, we additionally include an interaction term, $Online_{ic} \times Male_i$, in Equation (1) or Equation (2). The coefficient of this interaction term captures the incremental impact of online instruction on men relative to women, while the coefficient of the variable $Online_{ic}$ itself captures the impact on women. For robustness, we also split the sample by gender and estimate the effect of online instruction separately for men and women.

The key identification assumption for the estimation is that the exposure to online instruction is exogenous and uncorrelated to other factors in the error term that may affect learning outcomes. This is guaranteed through the experiment that allocates students randomly into different attendance groups based on the last digit of the student ID. To confirm that the randomization was successful, we conduct two balancing tests. First, we directly test whether student characteristics at baseline predict the proportion of online instruction. Table 2, Panel A shows that student baseline characteristics are not correlated with the share of online lectures. Second, we test whether student characteristics are balanced across the five attendance groups (Groups A to E). Table 2, Panel B, shows that attendance group dummies predict only one out of 12 student baseline characteristics (p-value = 0.096). Taken together, our balancing tests confirm that our randomization was successful.[7]

Table 2. Balance Tests

	(1)	(2)	(3)	(4)
Dependent Variable:	Effect of % Online Lectures		Joint Significance of Attendance Groups	
	Coefficient	S.E.	*F*-statistic	*p*-value
Male	-0.018	(0.029)	0.419	0.795
Age	-0.195	(0.136)	1.972	0.096
International Student	0.003	(0.020)	1.110	0.350
Business/Economics/Finance Majors	0.004	(0.031)	0.248	0.911
Informatics Majors	0.003	(0.020)	0.640	0.634
Other Majors	-0.007	(0.031)	0.303	0.876
Commute Time to University (min.)	1.164	(1.777)	0.533	0.711
High School Grade	-0.018	(0.029)	1.085	0.363
High School Self-study Intensity	-0.005	(0.011)	1.918	0.105
Group Study Preference	-0.050	(0.072)	0.875	0.478
Conscientiousness	-0.005	(0.055)	0.668	0.614
Extraversion	-0.063	(0.071)	0.757	0.554

Notes: Panel A tests whether the proportion of online lectures significantly predicts baseline characteristics. Each point estimate is derived from one OLS regression. In all regressions, we control for course fixed effects and cluster standard errors at the student level. The estimated coefficients and standard errors (S.E., in parentheses) are in Columns (1)–(2). Panel B tests whether students are randomly assigned into five attendance groups. Specifically, for each baseline characteristic, we test the joint significance of four teaching group dummies (one group omitted) in predicting the characteristic. The F-statistics and p-values derived from the tests are in Columns (3)–(4).

5 Results

[7] Appendix Table A.2 further tests whether the online proportion predicts baseline characteristics, separately for men and women. For female students, we find that none of the baseline characteristics are predicted by online instruction. For male students, only age significantly predicts the proportion of online lectures, which most likely represents a chance finding. Our results are robust to controlling for these characteristics (or student fixed effects).

5.1 Estimating the Impact of Online Instruction

Table 3 shows estimates for the impact of online instruction on the standardized exam score following Equation (1) or (2). We find that increasing exposure to online instruction has an overall negative effect on students' course performance. Column (1) uses the least restrictive specification controlling only for course fixed effects. We find that increasing the online proportion by 100 percentage points reduces the exam score by 0.103 SD (*p*-value = 0.043). Column (2) shows that this estimate basically stays the same when adding controls for student gender, age, international student status, commuting time, and major.

Table 3. The Impact of Online Instruction on Course Performance

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable: Standardized Exam Score					
% Online Lectures	-0.103**	-0.102**	-0.130**	-0.326***	-0.324***	-0.379***
	(0.051)	(0.050)	(0.053)	(0.080)	(0.079)	(0.085)
	[0.043]	[0.043]	[0.014]	[0.000]	[0.000]	[0.000]
% Online Lectures × Male				0.373***	0.369***	0.409***
				(0.103)	(0.102)	(0.108)
				[0.000]	[0.000]	[0.000]
Male		0.164***		-0.182**	-0.132	
		(0.034)		(0.090)	(0.089)	
International Student		-0.216**			-0.216***	
		(0.044)			(0.044)	
Age		-0.035***			-0.034***	
		(0.008)			(0.008)	
log(Commute Time)		-0.084***			-0.084***	
		(0.022)			(0.022)	
First Year in Study		-0.231***			-0.233***	
		(0.037)			(0.037)	
Observations	3,770	3,770	3,488	3,770	3,770	3,488
R-squared	0.519	0.557	0.838	0.523	0.557	0.839
Course Fixed Effects	✓	✓	✓	✓	✓	✓
Major Category Fixed Effects	-	✓	-	-	✓	-
Student Fixed Effects	-	-	✓	-	-	✓

Notes: The table shows the estimated impact of online instruction on students' course performance. Each column represents one Ordinary Least Squares (OLS) regression. The dependent variable is students' total exam score obtained for different courses, standardized across the whole sample to have a mean of 0 and a SD of 1. The key variable of interest is the proportion (%) of lectures assigned to be online. Standard errors in parentheses are clustered at the student level. p-values are in brackets. **p* <.1, ***p* <.05, ****p* <.01.

Column (3) uses our preferred and much more restrictive specification that controls for both course and student fixed effects. Including student fixed effects means that we are identifying effects of differential online exposure across courses within students—holding all observed and unobserved student-level factors constant. The inclusion of student fixed effects reduces our estimation sample by 7%—since we lose students observed in only one course—but increases the magnitude and precision of our estimate (p-value = 0.014). In this specification, we find that increasing the proportion of online lectures by 1 unit reduces exam score by 0.13 SD. This means that increasing the online share from 66% to 100%—the maximum effective variation in our data—reduces performance by 0.043 SD.

How large is this effect? Compared to previous experimental studies, our overall effect of 0.13 SD is more moderate than Kofoed et al. (2024) who find that going fully online lowered performance by 0.21 SD or Figlio, Rush and Yin (2013) who find an effect of 0.20 SD. We will provide a more detailed comparison to the existing literature in our meta-analysis in Section 7.

Effects by Gender: In Table 3, Columns (4)–(6), we additionally include an interaction term between the proportion of online lectures and student gender: "*% Online Lectures × Male*". The results show that the overall negative effect of online instruction is entirely driven by women. A 10 percentage points increase in online lectures reduces women's final exam score by 0.033 SD (see Column 4). This effect is very precisely estimated (p-value < 0.0001) and remains stable when including additional controls and student fixed effects (see Columns 5 and 6). The effect of online instruction for men is statistically significantly different from women and this gender gap is also very precisely estimated.

To make sure that the estimated gender gap is not confounded by potential interactions between gender and other control variables, we split the sample by gender and separately estimate the impact of online instruction for women and men in Table 4. By estimating separate models, we allow all coefficients, including the effects of covariates and the functional form of the relationship, to vary across genders. Columns (1)–(3) of Table 4 confirm that female students exposed to more online lectures perform significantly worse in the final exam. As the proportion of online lectures increases by 10 percentage points, their exam score decreases by 0.038 SD, using the specification including student fixed effects. This effect is precisely estimated; the 95% confidence interval of this effect is [−0.05, −0.02] SD.

By contrast, Columns (4)–(6) of Table 4 show that online instruction has no impact on male students' course performance. The estimated effect is close to zero. When controlling for student fixed effects, we find that increasing the proportion of online lectures by 10 percentage points is estimated to increase men's exam score by 0.003 SD, with the 95% confidence interval ranging from −0.001 to 0.015 SD. Table 4 shows that our conclusion do not depend on whether we cluster standard errors at the student level (in parentheses) or at the student-by-course level {in curly brackets}. Taken together, these results provide clear evidence that online instruction disproportionally harms women's performance relative to men.

Table 4. The Impact of Online Instruction by Gender

	(1)	(2)	(3)	(4)	(5)	(6)
	Female Students			Male Students		
% Online Lectures	-0.328***	-0.326***	-0.379***	0.049	0.048	0.030
	(0.079)	(0.078)	(0.084)	(0.066)	(0.065)	(0.068)
	{0.106}	{0.104}	{0.089}	{0.089}	{0.086}	{0.070}
Course Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	Yes	–	No	Yes	–
Student Fixed Effects	No	No	Yes	No	No	Yes
Observations	1,502	1,502	1,401	2,268	2,268	2,087
R-squared	0.549	0.572	0.841	0.506	0.550	0.837

Notes: The table shows the estimated impact of online instruction on the performance of women and men. The dependent variable is the total exam score standardized across the whole sample. In Columns (2) and (5), the specifications include the same set of student controls as in Column (3) of Table 3. Standard errors in parentheses are clustered at the student level; standard errors in curly brackets are clustered at the student-by-course level. **p* <.1, ***p* <.05, ****p* <.01.

Alternative performance measures: In addition to estimating effects for the total exam score—our most fine-grained performance measure—we also estimate effects for final course grades and the likelihood of passing the course. While these outcomes are ultimately functions of the exam score, they are much more salient to students and likely have a bigger influence on their self-perception and educational career. Appendix Table A.3 shows the effect of online instruction on course grade and the pass rate separately estimated for women and men. The results reveal a consistent pattern: Women assigned to more online lectures obtain lower final grades and become less likely to pass the course. Increasing the proportion of online lectures from 67% to 100%, lowers women's course grade by 0.15 SD, and their course passing rate by 0.07 percentage points—a 9.7% decline. Both of these effects are statistically significant at the 1% level. Again, we find no significant impact of online instruction on men's final grade or passing rate.

5.2 Heterogeneity by Other Student Attributes

Our results in Section 5.1 show large and significant gender differences in the response to online instruction. In addition to gender, we pre-specified eight other dimensions of heterogeneity in our pre-registration: previous achievement, nationality, self-study habits, group study preference, conscientiousness, extraversion, commute time, and major. For each of these characteristics, we divide students into two subsamples following our pre-registration and estimate heterogeneous effects using a fully interacted model.[8] Figure 2 shows estimates for the impact of online instruction on various sub-samples of students. With the exception of gender, none of the other subgroup differences reach statistical significance at the 5% level. The only other difference significant at the 10% level is between international students and students residing in Switzerland.

Multiple testing correction: Figure 2 raises the question whether the stark gender difference in response to online instruction could simply be a chance finding. To rule out that we are reporting a statistically significant, yet spurious gender gap, we need to account for multiple hypotheses testing. To adjust inference results accordingly, we apply the step-wise multiple testing approach of Romano and Wolf (2005) that controls family-wise error rate simultaneously to the 18 estimators of 9 models. Each model includes the online instruction variable and its interaction with the respective student attribute.[9]

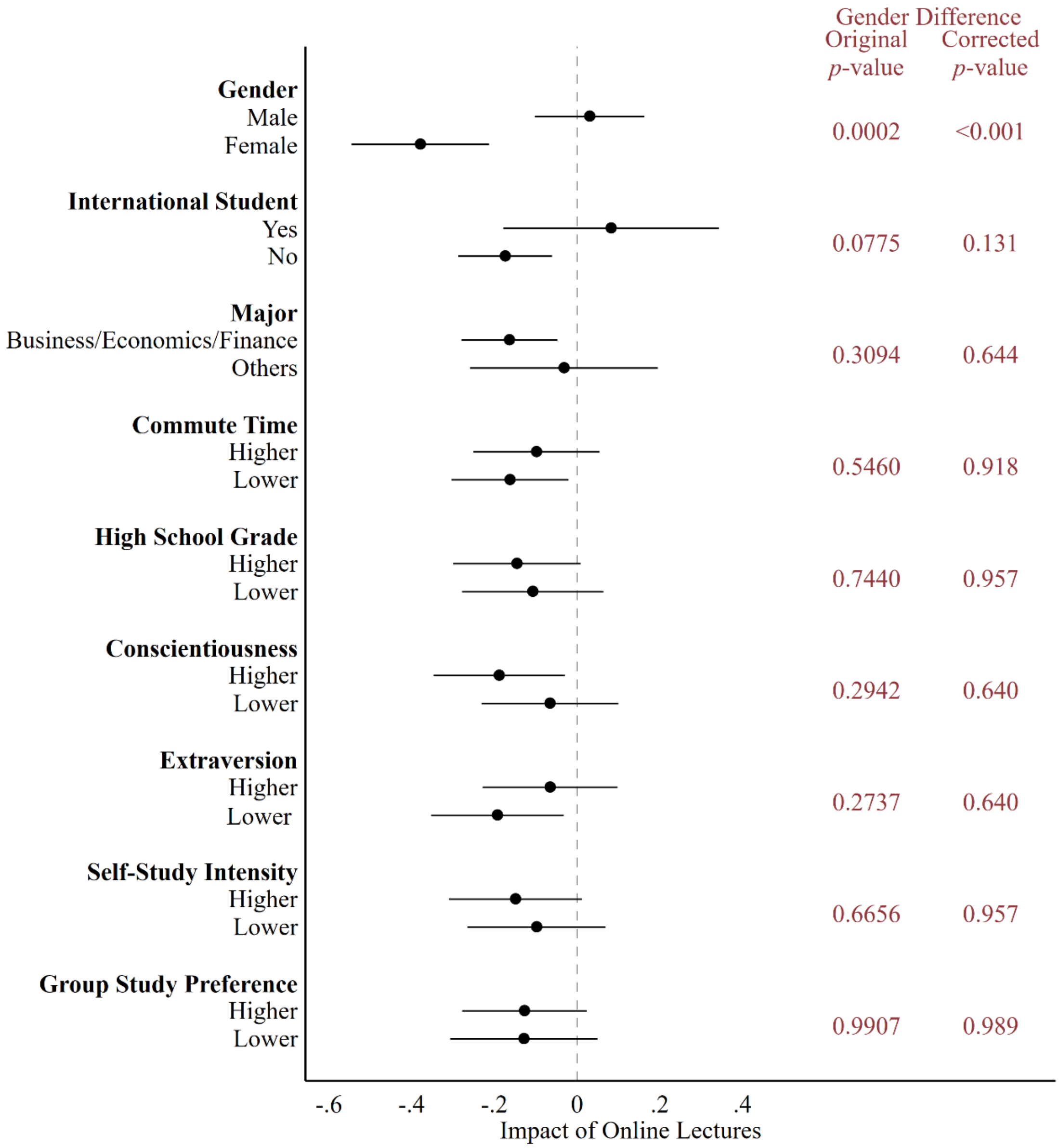
The last column of Figure 2 shows the corrected *p*-values for the subgroup differences.[10] While the gender gap in the effect of online instruction remains highly statistically significant after the correction (*p*-value = 0.001), none of the other differences across subgroups survive the correction. Notably, the gender gap even remains significant under the most conservative Bonferroni correction (*p*-value = 0.0036) where we multiply the original *p*-value with the number of conducted tests. Taken together, our results provide strong evidence that the gender difference in the effect of online instruction is robust and does not represent a false positive.

Figure 2. Heterogeneity of the Impact by Different Student Attributes

[8] Specifically, the model includes two interaction terms, between the online proportion and the two subsample indicators, always controlling for course fixed effects and student fixed effects.

[9] We apply the Romano and Wolf (2016) procedure by resampling 1,000 clusters of student observations, where each cluster consists of multiple observations from the same student, stratified by the five assigned attendance groups.

[10] Appendix Table A.4 shows the original 18 estimates and their corrected *p*-values.



Notes: The figure shows the estimated effects of online instruction on different sub-samples of students, divided based on their background characteristics. For continuous characteristics such as commute time and conscientiousness, the "Higher" group refers to students with above-median levels of the characteristic. For each characteristic, we use a model with two interaction terms (between online instruction and two subsample indicators), controlling for course fixed effects and student fixed effects. We cluster standard errors at the student level; the error bars indicate 95% confidence intervals. We then test the difference between the two interaction terms and report the *p*-values. Finally, we apply the Romano-Wolf multiple testing correction to all estimates and report the corrected the *p*-values for the between-group differences.

5.3 Impact on Earlier and Later Course Content

Our main results show that exposure to more online lectures during the first three weeks of the semester lowers women's course performance. To unpack how exactly online exposure affects exam performance at the end of the semester, we conduct an exam-question-level analysis. In this analysis, we differentiate between topics covered in earlier and later periods of the semester. To do that, we manually code how the 269 exam questions of the four courses in our sample relate to individual course lectures. Specifically, we acquire original teaching and exam materials and work with the course instructors to determine when the topic of each exam question was covered in a course. This allows us to map individual exam questions to one or multiple lectures. We then split all exam questions into three categories: (1) questions covered in weeks 1 to 3 of the semester—the period of our experiment; (2) questions covered in weeks 4 to 6—the period when restrictions on in-person attendance were relaxed; and (3) questions covered in weeks 7 to 13—the period when all instruction was conducted online. We then examine how the assigned proportion of online lectures in the first period affects student performance on questions in the three periods.

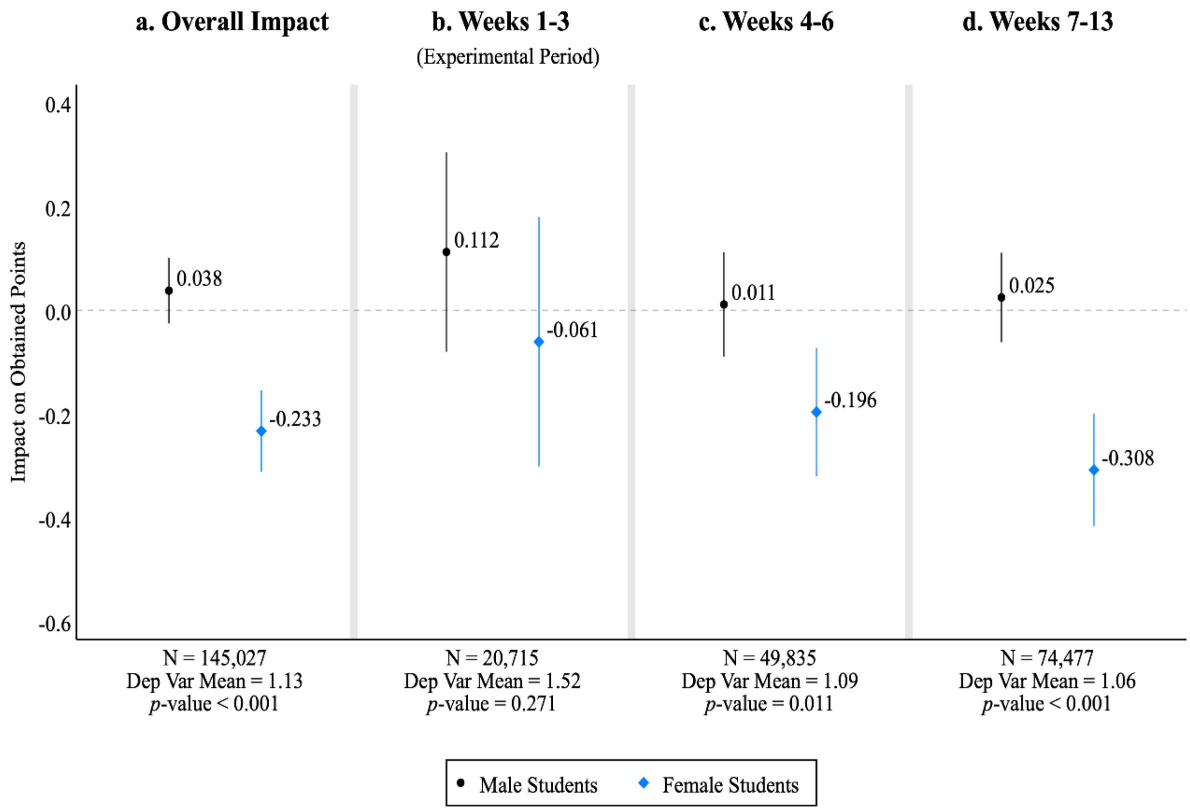
We conduct our analysis at the student-by-question level, using a total of 145,027 observations. The outcome variable is the number of points achieved for a question. As a robustness check, we also examine effects on a binary indicator of answering a question fully correctly. We estimate a model including two interaction terms ("% Online × Female" and "% Online × Male") as well as student fixed effects, and question fixed effects for each of the three time periods. We cluster standard errors at the student-by-question level.

Figure 3 plots the estimated effects of online instruction on exam points by teaching period and by gender. Panel (a) shows the overall effects pooling all periods, confirming our previous student-course level results. Panel (b) focuses on questions covered in the first three weeks of the semester and shows that online instruction in this period has no significant impact on the acquired question points of both men and women. [11] This is surprising, considering that this is the experimental period when differential online exposure was implemented. A plausible interpretation for the null effect could be that online instruction erodes the depth of initial understanding, which only reveals itself when students face integrated or advanced questions. It is also possible that questions about early content are too easy to detect differential online exposure effects.[12]

[11] We can also break down the treatment into student-by-question level variation: For each question, we define whether it was assigned online or in-person for a student. As Appendix Table A.11 shows, we find very similar effects when estimating how the question-level online assignment affects students' performance for different questions. This finding is broadly consistent with Cacault et al. (2001) who also use student-by-question level variation to show that students perform similarly for questions assigned with or without streaming access.

[12] While it may seem counterintuitive that online instruction during the first three weeks—the experimental period with randomized variation—did not significantly affect students' performance on the corresponding exam questions, several plausible explanations may account for this pattern. First, early course content likely focused on foundational

Figure 3. The Impact of Online Instruction on Obtained Points by Period



Notes: The figure plots the estimated effects of online instruction on obtained points for all exam questions (Panel a) and questions covered in weeks 1−3, 4−6, and 7−13, respectively. The obtained exam points vary from −1.5 to 4. The two gender-specific estimates for each period are estimated from one regression that uses student-question-level observations and includes two interactions terms: "% Online × Female" and "% Online × Male". We control for student fixed effects and question fixed effects in all regressions and cluster standard errors at the student-by-question level. The error bars indicate 95% confidence intervals; the *p*-value indicates the significance level of the gender difference between two estimates.

Panels (c) and (d) further show that early online exposure has a strong and precisely estimated negative effect on women's performance on questions covered in later periods of the semester. As the proportion of online lectures increases by 10 percentage points, women obtain 0.02 fewer points on questions covered in weeks 4 to 6 and 0.03 fewer points for questions on material covered in weeks 7 to 13. These effects correspond to approximately 1.8% and 2.9% of the average points obtained. Although online instruction generates strong dynamic spillover effects for women, we observe no significant spillover effects for men. The gender differences in the

or basic material, which may have been easier for students to grasp regardless of the instructional format, resulting in limited variation in performance (i.e., ceiling effects). Second, students often exhibit high motivation and engagement at the start of a semester, which could temporarily offset the disadvantages of online delivery. Third, the design and grading of early exam questions may not have been sensitive enough to capture subtle performance differences, particularly if they tested general prior knowledge rather than new content.

spillover estimates are statistically significant in period 2 (p-value = 0.011) and period 3 (p-value < 0.001).[13]

Taken together, we find that online instruction does not affect women's performance on exam questions related to early lecture content, which primarily covers basic foundational concepts. However, women who experienced more online instruction perform notably worse on exam questions covering content taught later in the semester, which involves more complex topics building upon earlier content. These negative dynamic spillover effects might result from either (1) women's incomplete understanding of fundamental concepts when taught online, or (2) lower motivation and effort in later learning stages. Overall, these findings foreshadow that online instruction could be especially detrimental to women's long-term educational outcomes.

5.4 Course-Level Heterogeneity

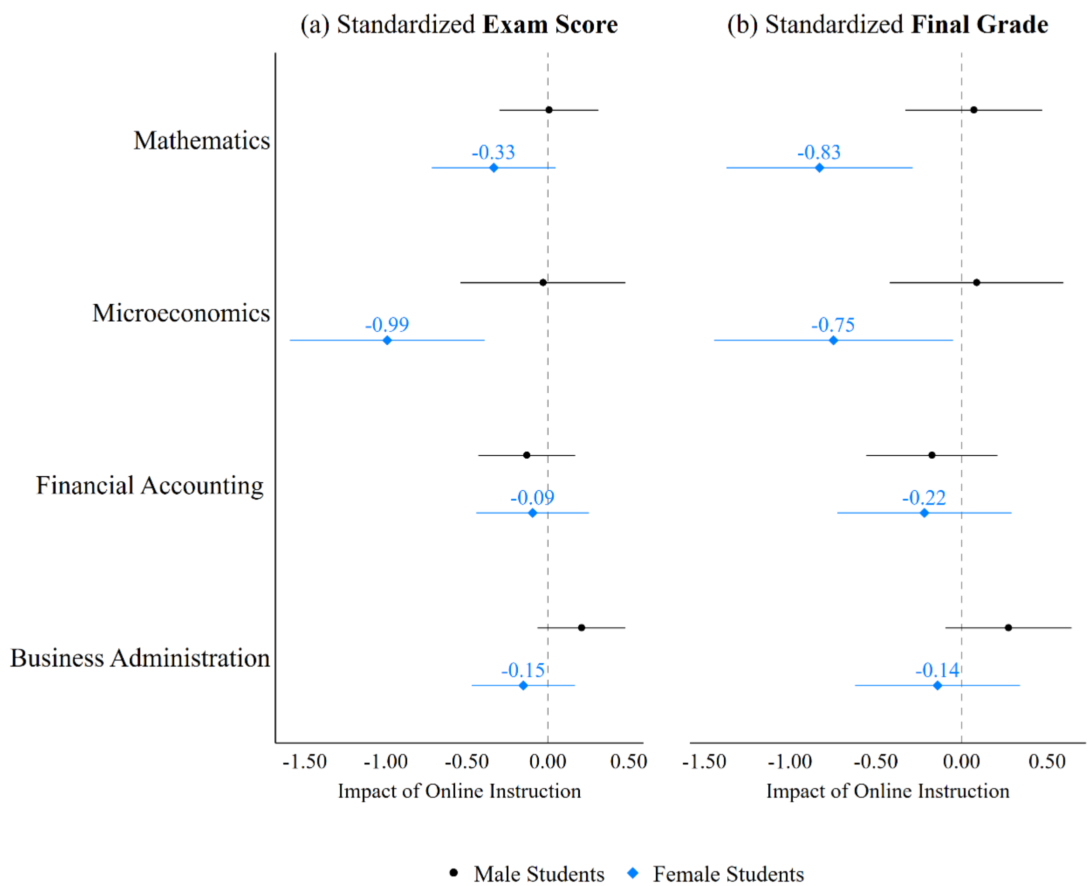
After investigating student and question level heterogeneity, we next examine whether the impact of online exposure differs across the four courses in our data. Figure 4 plots the estimated impact of online exposure in Mathematics, Microeconomics, Financial Accounting, and Business Administration, on the obtained exam score (Panel a) and final grade (Panel b). Figure 4 highlights that the overall negative effect for women is almost completely driven by the Mathematics and Microeconomics courses and not statistically significant in Financial Accounting and Business Administration. A 10 percentage points' increase in the proportion of online lectures in Microeconomics reduces women's exam score for the course by 0.1 SD and their final grade by 0.08 SD. Similarly, a 10 percentage points' increase in online Mathematics lectures reduces women's exam score by 0.03 SD and final grade by 0.08 SD.

Why is online instruction particularly harmful for women's performance in the relatively math-intensive courses? Although we cannot pin down why the effect varies across courses, two explanations appear plausible. First, given the results from the previous section and the cumulative nature of math-intensive courses, women may have a particularly difficult time catching up after initial online exposure. Any learning gaps or misunderstandings arising from the initial online

[13] In Appendix Table A.6, Panel A, we show the full regression results for the estimates in Figure 3. Panel B further presents estimated effects on the binary outcome of correctly answering an exam question. The results remain consistent: While online instruction does not significantly affect men's correct rate for questions in any period or women's correct rate for questions covered in the first three weeks, it significantly reduces women's likelihood of correctly answering questions covered in later periods of the semester.

instruction may disproportionately accumulate in math-intensive courses, making subsequent topics increasingly challenging.[14] Second, it is also possible that because women feel less confident about mathematical subjects, they are more sensitive to reduced social interaction in the online learning environment. The lack of immediate feedback, clarification opportunities, and peer support could be particularly harmful for subjects where women are less self-confident than men.[15]

Figure 4: The Impact of Online Instruction on Performance by Course



Notes: The figure shows the estimated impact of online instruction on students' performance in four courses. The outcome variable is the standardized exam score in Panel (a) and the standardized course grade in Panel (b). The results for each gender group in each panel are estimated with one specification that includes four interaction terms between the online proportion variable and four course indicators. All regressions control for both course fixed effects

[14] Indeed, when analyzing the impact of online instruction by the teaching period of different exam questions, we find that Mathematics and Microeconomics display strong dynamic spillover effects (as in Figure 3)—meaning that early online lectures in these courses significantly lower women's performance for later lecture content. By contrast, for Financial Accounting and Business Administration, we find limited evidence of dynamic spillover effects.

[15] We cannot rule out the explanation that the instructors of Microeconomics and Mathematics are more engaging, so that women benefit more from the in-person learning experience in these courses. The course-level heterogeneity cannot simply be explained by the instructor's gender, because only the math course was taught by a female professor.

and student fixed effects and cluster standard errors at the student level. The error bars indicate 95% confidence intervals. See Appendix Table A.10 for complete regression results.

5.5 The Longer-Term Impact on Educational Attainment

Does exposure to online instruction at the beginning of university studies—a crucial period of learning—also influence students' educational outcomes in the longer term? Based on our results in Section 5.3 documenting dynamic spillover effects, it appears possible that learning losses in the first semester can accumulate and adversely affect longer-run outcomes.

We have so far exploited student-by-course level variation to examine the short-term impact of online instruction on course performance. This strategy does not allow us to analyze effects on longer-term outcomes, which vary at the student but not at the course level. We also cannot use the average proportion of online lectures in four courses, because our "partial attendance" policy was designed to provide equal opportunities for in-person instruction for all students and minimized the overall exposure to online instruction at the student level. Guided by our finding in Section 5.4 that the impact of online instruction is very much concentrated in the two math-intensive courses, we examine how online lectures in these courses impact longer- term student attainment.

Using administrative data, we construct four longer-term outcomes: (1) the share of course credits earned in the 2020/21 academic year (i.e., the quotient of earned credits by booked credits); (2) average course grades in the 2020/21 academic year; (3) dropping out of the original study program within one year after the experiment; and (4) dropping out within two years after the experiment.

Table 5 shows how the share of online lectures in Mathematics and Microeconomics affects the longer-term educational outcomes, restricting the analysis to students that we observe in both courses. The results in Columns (1)–(2) show that increasing exposure to online instruction in math-intensive courses reduces women's overall performance during the 2020/21 academic year. As the proportion of online lectures increases by 10 percentage points, women's obtained share of course credits decreases by 4.9 percentage points, and their average course grade decreases by 0.09 points—2.2% of the mean. For men, in contrast, online instruction in these courses seems to have a positive effect on their performance during the academic year—the gender gap in the impact of online instruction is statistically significant at the 5% or 10% level depending on the outcome.

In Columns (3)–(6), we restrict our analysis to "on-track students," who were in their first study year at the start of our experiment. Results remain very similar when focusing on this subset of 735 freshmen students. Columns (5)–(6) indicate that greater exposure to online lectures in math-intensive courses significantly increases dropout risk within the first two years of the undergraduate program. A 10 percentage points increase in the proportion of online lectures raises the likelihood that women drop out of their original study program within two years by 7.3 percentage points—a 19% increase. For men, in contrast, the impact goes in the opposite direction. Taken together, our results highlight that online instruction has a lasting influence on students' academic trajectory. Learning losses due to increased exposure to online lectures in the first semester—a crucial period for laying the huma capital foundation—appear to accumulate and adversely affect women's longer-run outcomes.

Table 5. The Long-Term Impact of Online Instruction in Math-Intensive Courses

	(1)	(2)	(3)	(4)	(5)	(6)
	All Students		First-Year Students			
	%Credits 2020/21	Avg Grade 2020/21	%Credits 2020/21	Avg Grade 2020/21	Dropout in Year 1	Dropout in Years 2
% Online Lectures in Math/Micro	-0.493**	-0.928*	-0.504**	-0.965*	0.731**	0.733**
	(0.208)	(0.533)	(0.231)	(0.580)	(0.333)	(0.367)
% Online Lectures in Math/Micro × Male	0.659**	1.187*	0.683**	1.299*	-0.867**	-1.203***
	(0.269)	(0.717)	(0.291)	(0.773)	(0.400)	(0.457)
Observations	836	836	735	735	735	735
R-squared	0.032	0.031	0.028	0.026	0.021	0.042
Dep Var Mean	0.706	4.211	0.714	4.247	0.211	0.382

Notes: The table shows the impact of online instruction in Microeconomics and Mathematics on students' longer-term educational outcomes. The variable "% Online Lectures in Math/Micro" refers to the average proportion of online lectures in Mathematics and Microeconomics, which varies from 67% to 92%. The dependent variables in Columns (1)–(4) are the share of credits earned ("%Credits") in the 2020/21 academic year and the average final grade ("Avg Grade") obtained for courses taken during the 2020/21 academic year, respectively. The dependent variables in Columns (5) and (6) are binary indicators for dropping out of the original study program at the business school by fall 2021/22 and fall 2022/23. The analyses in Columns (1)–(2) include all students who were enrolled in the Mathematics and Microeconomics courses during the experiment; the analyses in Columns (3)–(6) focus on students who were in their first-semester study during the experiment. All specifications use OLS regressions or linear probability models, in which we control for gender and major categories. Robust standard errors are in parentheses. *p <.1, **p <.05, ***p <.01.

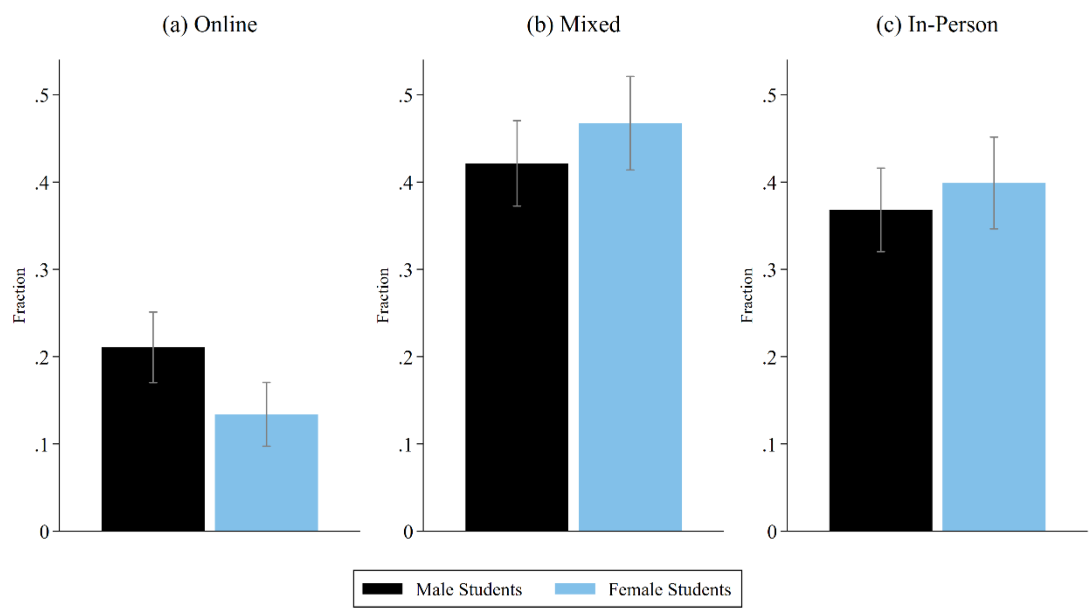
6 Possible Mechanisms

Our results so far establish a clear gender gap in the effects of online instruction on learning outcomes. The effects are driven by relatively math-intensive courses and appear to translate into lasting changes in education outcomes. In this section, we explore why female students do worse when experiencing lectures online, while men seem to be less sensitive to the teaching mode. We first examine whether men and women differ in their preference for online instruction in Section 6.1. We then examine how online instruction affects the overall course attendance in Section 6.2. Finally, we analyze how online instruction affects self-reported learning experiences of men and women in Section 6.3.

6.1 Gender Differences in Preferences

Self-reported preference: We start by providing descriptive evidence on gender differences in students' preferred instruction mode. In our endline survey, conducted at the end of the first semester, we asked students: "What kind of lectures would you personally prefer for the future? (Provided that the general Covid-19 situation allows this.)" and provided three answering options: (a) online lectures; (b) mixed/partial attendance; and (c) in-person lectures. As Figure 5 shows, 13% of women and 21% of men report online as their preferred teaching mode. Women are thus 36% less likely than men to choose online lectures (instead of mixed or in-person lectures) as their preferred teaching mode. This gender gap is statistically significant at the 1% level.

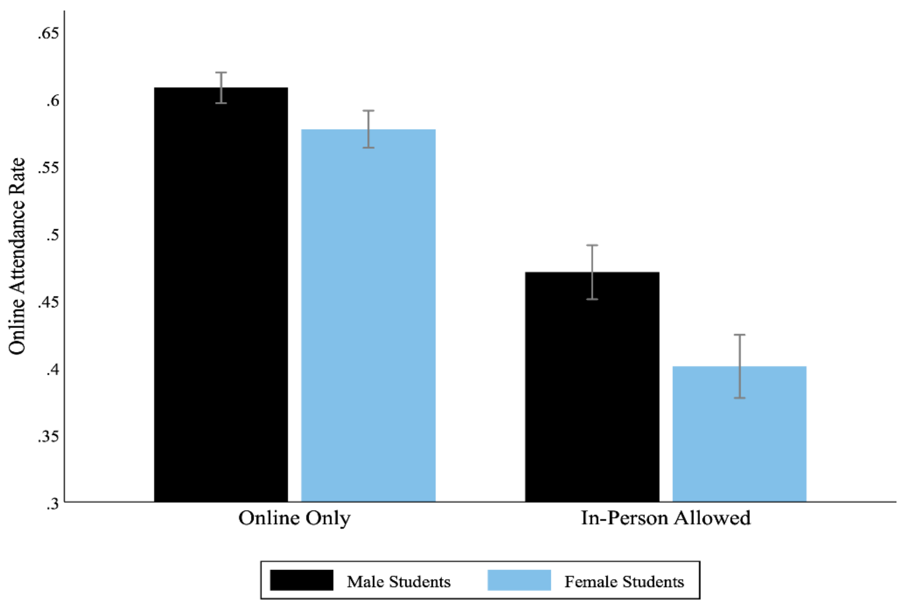
Figure 5. Preference for Different Teaching Modes by Gender



Notes: The figure shows the fraction of male and female students who prefer each type of instruction mode, with 95% confidence intervals. The question is: "*What kind of lectures would you personally prefer for the future? (Provided that the general Covid-19 situation allows this.)*" The three options are online lectures, mixed/partial attendance, and in-person lectures. The gender difference in Panel (a) is statistically significant (p-value = 0.007). In Panels (b) and (c), the gender difference is not significant.

Revealed preference: After documenting a gender gap in students' stated preferences for online instruction, we next look at students' revealed preference using zoom attendance data for the Microeconomics course. We use this data to document how students' online attendance varies with their assigned teaching mode. Figure 6 shows the attendance rates by gender in two conditions: when students were assigned to attend a lecture online ("Online Only") and when in-person attendance is allowed ("In-Person Allowed"). Figure 6 reveals two facts: First, irrespective of the assigned lecture mode, male students are more likely than female students to attend online lectures. Second, this gender online attendance gap is much larger when in-person attendance is allowed. When assigned to an in-person lecture, women are 7 percentage points (15%) less likely than men to attend the lecture on Zoom—suggesting that women took up the opportunity to physically show up in the class. However, when assigned to an online lecture, women are only 3 percentage points (5%) less likely than men to attend the lecture on Zoom. This 4-percentage-points difference in the gender gap suggests that women are more likely to attend lectures in person (instead of online) whenever in-person attendance is possible.

Figure 6. Online Attendance Rate by Gender and Assigned Teaching Mode

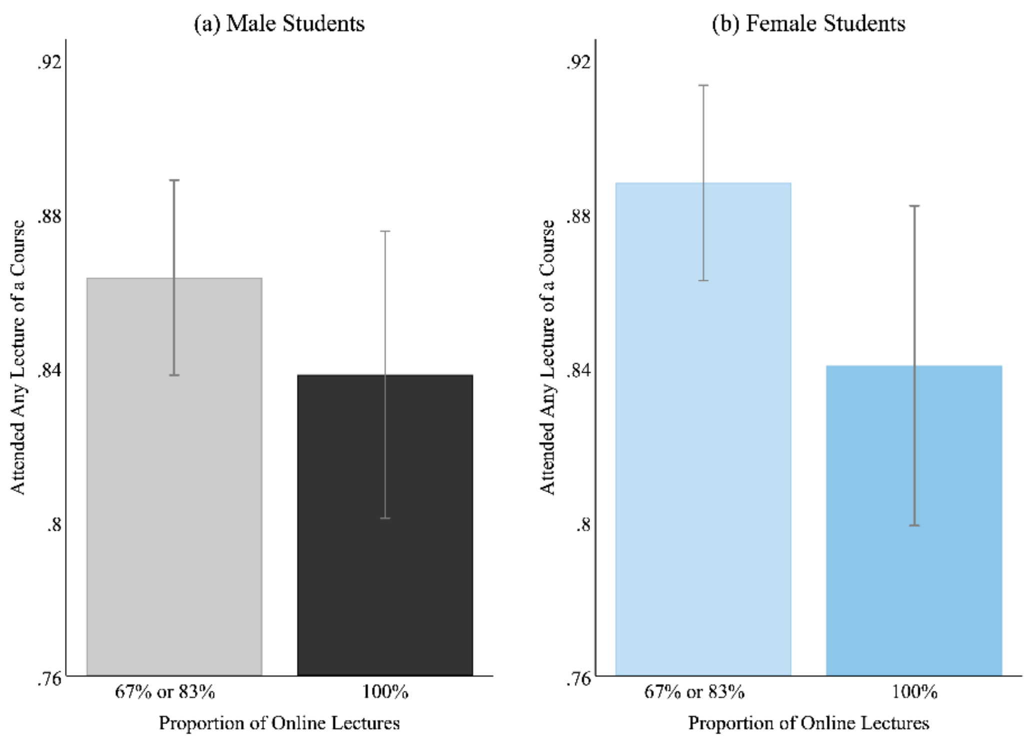


Notes: The figure shows the fraction of male and female students attending online Zoom lectures in the Microeconomics course, depending on whether in-person attendance is allowed for a lecture. The sample includes 15,890 student-lecture-level observations across 7 weeks. Error bars indicate 95% confidence intervals. The gender differences in both panels are statistically significant (p-value < 0.001).

6.2 Overall Course Attendance

Figures 5 and 6 above provide consistent evidence that women hold a stronger preference than men for in-person instruction. This preference for in-person lectures may affect women's overall course attendance rate—those who are assigned to more online lectures may end up attending fewer lectures. To directly test this, we asked students in the endline survey whether they attended any lectures during the semester, for *each* course they are enrolled in. 79% of students report that they attended at least one lecture of all the courses; 10% of students did not attend any lectures of any course; and the other 11% attended at least one lecture of some of the courses.

Figure 7. Overall Course Attendance by Gender and Online Proportion



Notes: The figure plots the likelihood of having attended any lecture of a course by gender and assigned proportion of online lectures in the course. The attendance rate is reported in the endline survey. Error bars indicate 95% confidence intervals. The difference in attendance rate by online proportion is insignificant in Panel (a) but statistically significant in Panel (b), with a *p*-value of 0.045.

Figure 7 shows how the extensive-margin course attendance rate varies with the assigned proportion of online lectures in a course, separately for men and women. We find that when assigned 100% of online lectures at the start of the semester, men and women have a similar course attendance rate (approximately 84%). When assigned fewer online lectures (i.e., more in-person lectures), women's course attendance rate increases to 89%. This attendance gap by online exposure is statistically significant (p-value = 0.045) and becomes more precisely estimated when controlling for course fixed effects and student fixed effects (p-value = 0.034). For men, the attendance gap is much smaller and does not reach statistical significance. This finding suggests that one channel through which online instruction reduces women's exam performance is that it discourages them from attending the lectures where they learn exam-relevant course content.

6.3 Learning Experiences

Our results in the previous section highlight that women prefer in-person over online instruction, and women assigned more online lectures have a lower course attendance rate. To better understand why women dislike online instruction, we asked students to evaluate different aspects of their learning experiences related to lectures, professors, and peers—for each course—in the endline survey. Then, we estimate how the proportion of online lectures affects students' evaluations of these course-specific aspects.

Evaluations of courses and lectures: We use three questions to measure students' satisfaction with different courses: "How much did you like [course name]?" "The lectures helped me to better understand the course content;" and "The lectures were clearly structured and easy to follow." Students answer the first question on a Likert scale from 1 (very poor) to 10 (very good) and the latter two questions on a scale from 1 (strongly disagree) to 10 (strongly agree).

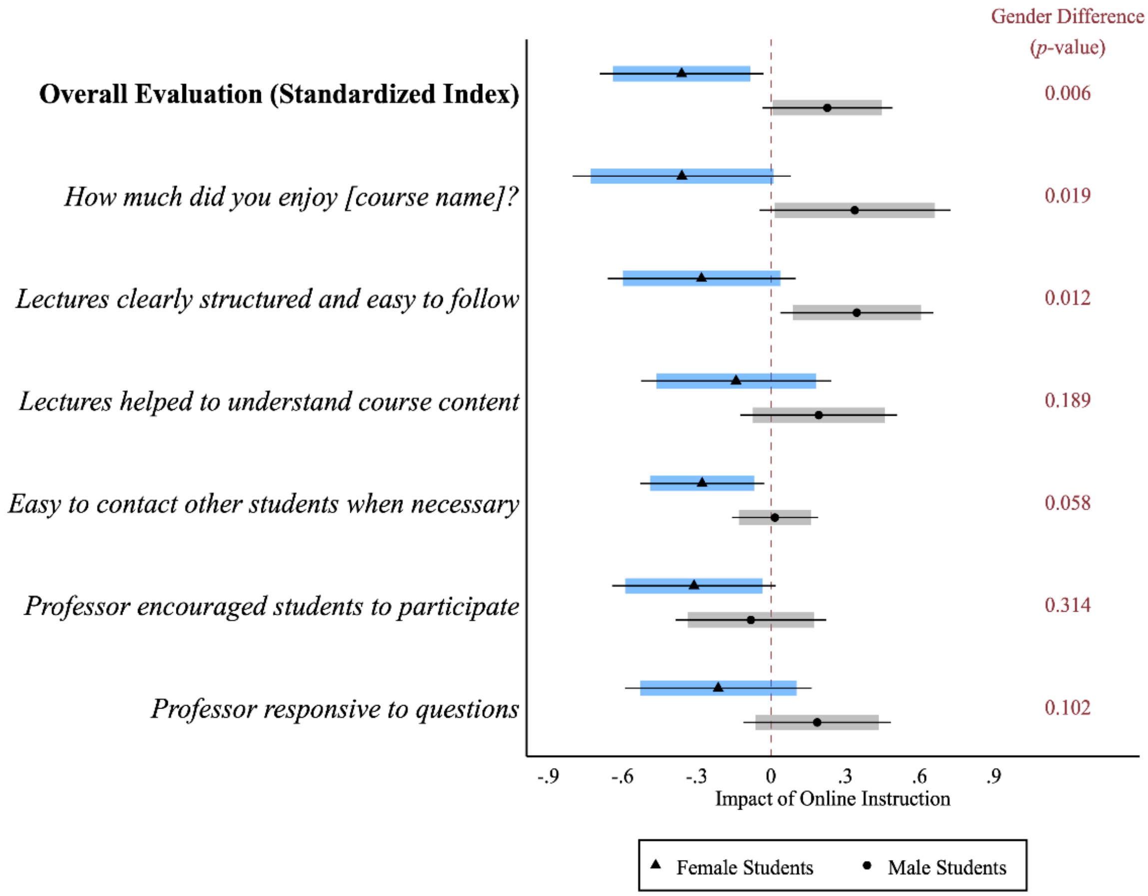
Evaluations of professors and peers: Another feature of lectures that may be crucial for learning is whether they are engaging enough and whether students can readily get support from the teacher and peers. We measure this feature using students' agreement with the following statements: "The professor regularly encouraged students to actively participate in the lecture;" "The professor has been responsive to student questions;" and "Regarding the course, it was easy to contact other students when necessary" all on a scale of 1 (strongly disagree) to 10 (strongly agree).

Based on these measures, we construct a standardized index that combines all dimensions of learning experiences, with larger values representing more-positive evaluations.[16] The first two estimates in Figure 8 show the effects of online instruction on the overall evaluations of men and

[16] Specifically, we standardize each measure to have a mean of 0 and a SD of 1. Then, we derive the mean of these standardized measures. Finally, we standardize the mean to get the overall evaluation index.

women. We find that increasing the proportion of online lectures by 10 percentage points decreases women's overall evaluation by 0.036 SD (p-value = 0.032) but improves men's overall evaluation by 0.022 SD (p-value = 0.09). The gender gap in the impact on learning experiences is statistically significant at the 1% level (p-value = 0.019). When breaking down this overall impact, the estimates paint a clear picture that women are unsatisfied with almost *all* aspects of online lectures: Women exposed to increased online instruction find a course less enjoyable, consider the lectures less clear and less helpful, experience more difficulty in contacting peers, and evaluate the professors as less engaging and less responsive.

Figure 8. The Impact of Online Instruction on Learning Experiences



Notes: The figure shows the estimated effects of online instruction on a standardized index for overall course evaluations and specific dimensions of the index. The specific outcomes are measured with the following questions in the endline survey: (1) "How much did you like [course name]?" (2) "The lectures were clearly structured and easy to follow;" (3) "The lectures helped me to better understand the course content;" (4) and "Regarding the course, it was easy to contact other students when necessary;" (5) "The professor regularly encouraged students to actively participate in the lecture;" and (6) "The professor has been responsive to student questions." Students answer all these questions on a 10-point Likert scale, with higher values indicating a higher degree of enjoyment or agreement. For each outcome, we use a model including two interaction terms "%Online × Female" and "%Online × Male" to estimate the impact of online instruction on men and women. We then test the difference between the two interaction terms and report the *p*-value. Each regression controls for student fixed effects and course fixed effects and cluster standard errors at the student level. Error bars indicate 90% and 95% confidence intervals.

Taken together, the results in this section show that compared to men, women are less satisfied with online lectures, possibly because they consider online lectures as less informative and less interactive. As documented in prior work, the "gender digital divide" implies that women may be less comfortable with using digital tools for learning purposes, compared to men. Women may also have a stronger demand for social interactions in the learning process. As a result, the lack of engagement and peer support in online learning environments may also hamper their learning outcomes.

7 Meta-Analysis

To better place our results into the existing literature, we conduct a simple meta-analysis on the impact of online instruction. For our meta-analysis, we searched Google Scholar and Research Rabbit using combinations of the search terms "online education", "online instruction", "gender", "in-person education", "in-person instruction", "live instruction", "live education", "effect", "university". We restricted the selection to studies at the tertiary education level which identified the causal effect of being exposed to higher education through experimental or quasi-experimental designs (RCT or IV). We then cross-checked the bibliography of the selected papers to make sure we did not overlook any relevant paper. Based on the procedure we identified 10 studies that estimate the effects of online instruction on exam performance in higher education.[17] Appendix Table A.12 provides the details of these 10 articles, including their setting, identification strategy, treatment vs. control design, and sample. To be precise, seven of these 10 studies estimate the impact of online instruction compared to in-person instruction. The remaining three studies estimate the effect of more vs. less in-person instruction (Joyce et al., 2015), access to online streaming (Cacault et al., 2021), or hybrid vs. in-person instruction (Bowen et al., 2014). In terms

[17] Outside of higher education, there are also a handful of studies that examine the impact of school closure or remote learning in schools, including Lichand et al. (2022), Bacher-Hicks et al. (2021), and Musaddiq et al. (2022). Lichand et al. (2022) find that the shift to remote learning in Brazil substantially increased dropout rates and reduced test scores, especially among disadvantaged and female students. Bacher-Hicks et al. (2021) document large disparities in online learning engagement across U.S. school districts, with drastic increases in search intensity for online resources in higher-income areas. Musaddiq et al. (2022) show that the pandemic triggered significant shifts in school enrollment patterns in the U.S., with public school enrollment declining, particularly in kindergarten, while homeschooling and private school enrollment rose.

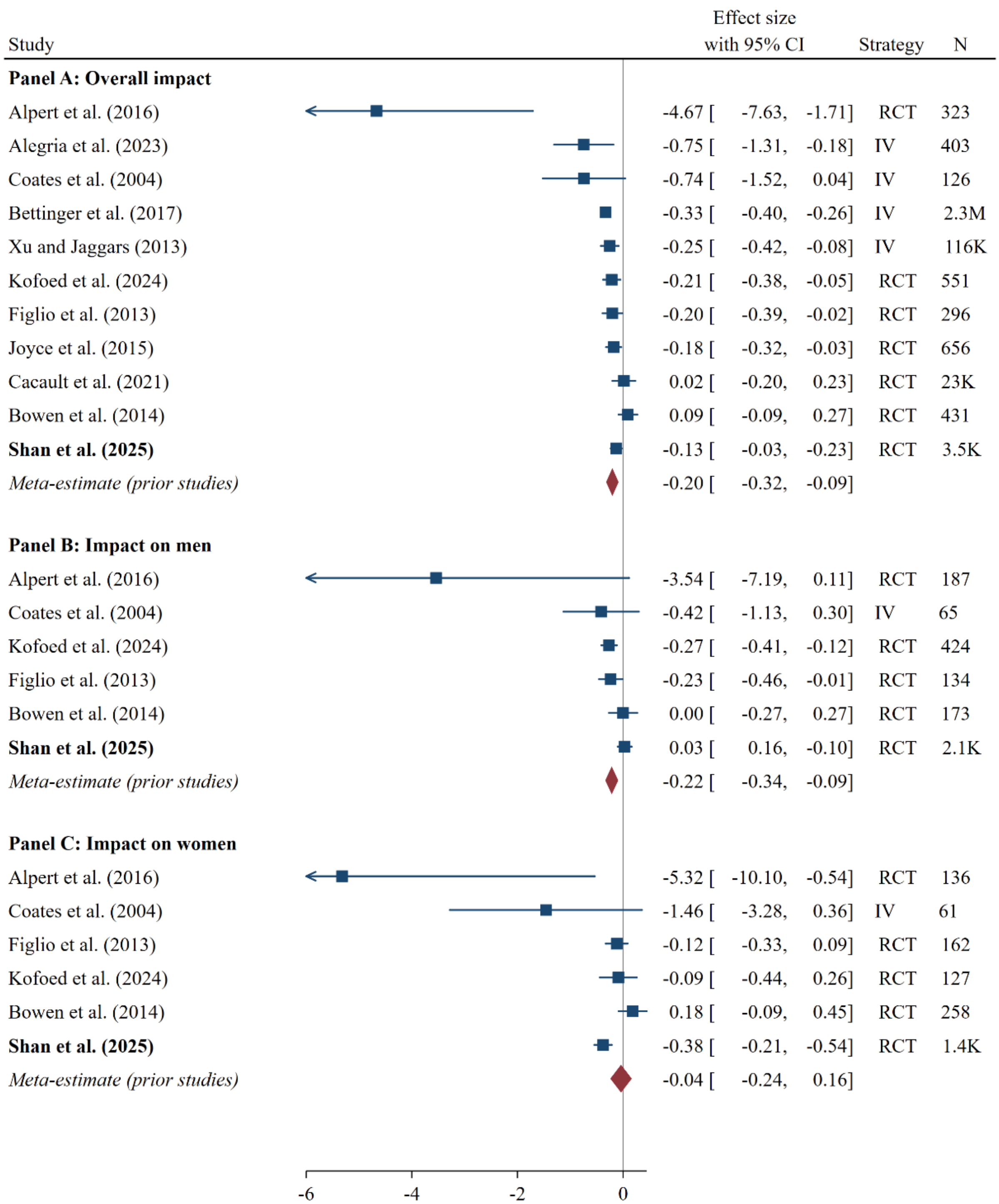
of the empirical strategy, six out of the 10 studies conduct randomized experiments, and the other four studies use instrumental variable (IV) approaches. Considering the sample size, the median number of observations in prior studies is 491, highlighting that the majority of previous studies in this literature are underpowered.

Although all studies report estimates for the overall impact of online (or hybrid) instruction, we do not observe gender-specific estimates for most studies. Only three out of the 10 prior studies (Figlio et al., 2013; Bowen et al., 2014; Kofoed et al., 2024) report results by gender in the main paper or online appendix. We obtained gender-specific results for two additional studies by either using the authors' replication package (Alpert et al., 2016) or successfully contacting authors to produce the results for us (Coates et al., 2004). Thus, we have gender-specific estimates for five studies in total. To keep the estimates comparable across studies, we ensure all point estimates and standard errors are measured in standard deviations of the outcome of each study. We do this by dividing estimates and standard errors by the standard deviation of the outcome in all studies that did not report their estimates in standardized units.

Figure 8, Panel A, shows estimates of the overall impact of online education, including the effect value, 95% confidence intervals, and sample size. Eight out of 10 studies report negative point estimates of the impact of online instruction. The meta-estimate of the average treatment effect is -0.204 SD, with a standard error of 0.059 (*p*-value = 0.00061) and significant between-study variations (*p*-value = 0.00001). These results highlight that (1) the existing literature largely agrees that the overall impact of online instruction is negative and that (2) the average treatment effect of -0.204 SD in prior studies is more sizable than our overall estimate of -0.13 SD, which is not included in the meta-estimate.

Panels B and C of Figure 8 show estimates of the impact split by gender. Panel B provides estimates for men. The meta-estimate of the average online effect is -0.215 SD with a standard error of 0.062 (*p*-value = 0.00052). Panel C summarizes estimates for women. The meta-estimate of the average online effect for women is -0.039 SD with a standard error of 0.103 (*p*-value = 0.703). While the meta-estimate for women is substantial in magnitude, it is not statistically significant. Comparing the meta-estimates for women and men highlights that the existing literature as a whole has not documented a gender difference in the effect of online instruction.

Figure 8: Meta-analysis on the Impact of Online instruction



Note: All estimates are in the unit of standard deviation of the outcome variable; the error bars indicate 95% confidence intervals. The meta-analysis does not include our own paper (weight = 0). Panel A exhibits the effect on the full sample; in Panel B and C the sample is restricted to men and women, respectively. The random-effects restricted maximum likelihood (REML) estimate for the overall sample is -0.204 SD (*p*-value = 0.00061). There is substantial between-study heterogeneity (τ^2 = 0.02, I^2 = 70.34%, H^2 = 3.37), as confirmed by the Cochran's Q test (Q(9)=38.9, p= 0.000012). The REML estimate for the men sample is -0.215 SD (*p*-value = 0.00052). There is low between-study heterogeneity (τ^2 = 0.0018, I^2 = 7.95%, H^2 = 1.09), as confirmed by the Cochran's Q test (Q(4)=6.46, p= 0.167). The REML estimate for the women sample is -0.039 SD (*p*-value = 0.703). There is medium between-study heterogeneity

(τ^2 = 0.01, I^2 = 27.06%, H^2 = 1.37), as confirmed by the Cochran's Q test (Q(4)=10.32, p= 0.035). A test of subgroup differences indicates significant variation between the three subgroups ($Q_b(2)$=2.33, p=0.311).

Figure 8 highlights that our study provides the first evidence of how online instruction differentially affects men and women. Our study is also the first that documents a precisely estimated negative effect of online instruction on women. While the results by Alpert et al. (2016) also suggest a negative impact for women, this effect is very nosily estimated and was not reported in the original study (we derived the estimates using their replication package). Although there are many factors that may affect the gender gap in the impact of online instruction, it is worth highlighting that the lack of statistical power can be an important reason for why the literature has remained silent on online education gender gap so far.

Taken together, our meta-analysis situates our findings within the broader literature and provides two key insights: First, our overall negative point estimate is consistent with most prior research. Second, our main finding that women experience disproportionately negative impacts from online education has remained overlooked. Our precisely estimated gender-specific results, highlight that online instruction seems to harm women in male-dominated, math-intensive learning environments.

8 Conclusion

This paper provides the first experimental evidence that online instruction in higher education disproportionately harms women's academic outcomes. In a large-scale field experiment that varied students' exposure to online and in-person lectures, we find that increased online instruction significantly reduces women's exam performance but leaves men unaffected. The adverse effect on women is especially pronounced in math-intensive courses.

Exploring the underlying mechanisms, we document gender disparities in learning preferences and experiences. When assigned to more online lectures, women report lower course satisfaction, experience greater challenges in contacting their peers, and perceive their instructors as less engaging—these effects do not apply to men. The stronger preference for face-to-face learning also reduces women's overall lecture attendance when in-person attendance is limited.

Concerningly, the negative impacts of online instruction are not limited to contemporaneous outcomes but also extend to women's educational trajectories. Exposure to online instruction during the first weeks of their studies reduces women's credit accumulation, depresses their average grades in the following academic year, and increases their dropout rates. These results underscore how even gradual and short-lived shifts toward online instruction can produce lasting negative impact on women's educational careers. Policymakers and educators must be aware of this risk when expanding seemingly cost-effective online education. Based on our results, improving women's engagement, peer-to-peer contact, and instructor accessibility appear to be the most promising targets for mitigating the unintended consequences of online instruction.

References

Adams-Prassl, Abi, Kotaro Hara, Kristy Milland, and Chris Callison-Burch. 2023. "The gender wage gap in an online labor market: The cost of interruptions." *Review of Economics and Statistics*, 1–23.

Alegría, Rodrigo, Pablo Cárabe, Alejandro Chahoud, and Ainara González de San Román. 2023. "The novel liquid learning system and the online gap in academic performance." *Education Economics*, 31(5): 599–616.

Alpert, William T, Kenneth A Couch, and Oskar R Harmon. 2016. "A randomized assessment of online learning." *American Economic Review*, 106(5): 378–382.

Anderson, Michael L. 2008. "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American statistical Association*, 103(484): 1481–1495.

Arbaugh, J Ben. 2000. "Virtual Classroom Characteristics and Student Satisfaction with Internet-Based MBA Courses." *Journal of Management Education*, 24(1): 32–54.

Bacher-Hicks, Andrew, Joshua Goodman, and Christine Mulhern. 2021. "Inequality in household adaptation to schooling shocks: Covid-induced online learning engagement in real time." *Journal of Public Economics*, 193: 104345.

Belenky, Mary F, Blythe M Clinchy, Nancy R Goldberger, and Jill M Tarule. 1988. "Women's ways of knowing: The development of self, voice, and mind." *Curriculum Inquiry*, 18(1): 113–121.

Bettinger, Eric P, Lindsay Fox, Susanna Loeb, and Eric S Taylor. 2017. "Virtual classrooms: How online college courses affect student success." *American Economic Review*, 107(9): 2855–2875.

Bowen, William G, Matthew M Chingos, Kelly A Lack, and Thomas I Nygren. 2014. "Interactive learning online at public universities: Evidence from a six-campus randomized trial." *Journal of Policy Analysis and Management*, 33(1): 94–111.

Cacault, M Paula, Christian Hildebrand, Jérémy Laurent-Lucchetti, and Michele Pellizzari. 2021. "Distance learning in higher education: evidence from a randomized experiment." *Journal of the European Economic Association*, 19(4): 2322–2372.

Carvajal, Daniel, Catalina Franco, and Siri Isaksson. 2024. "Will Artificial Intelligence Get in the Way of Achieving Gender Equality?" *Working Paper*.

Coates, Dennis, Brad R Humphreys, John Kane, and Michelle A Vachris. 2004. "No significant distance" between face-to-face and online instruction: Evidence from principles of economics." *Economics of Education Review*, 23(5): 533–546.

Cook, Cody, Rebecca Diamond, Jonathan V Hall, John A List, and Paul Oyer. 2021. "The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers." *The Review of Economic Studies*, 88(5): 2210–2238.

Deming, David J, Claudia Goldin, Lawrence F Katz, and Noam Yuchtman. 2015. "Can online learning bend the higher education cost curve?" *American Economic Review*, 105(5): 496–

501.
Else-Quest, Nicole M., Janet Shibley Hyde, and Marcia C. Linn. 2010. "Cross-national patterns of gender differences in mathematics: a meta-analysis." *Psychological bulletin* 136(1): 103-127.
Emanuel, Natalia, Emma Harrington, and Amanda Pallais. 2023. “The power of proximity to coworkers: training for tomorrow or productivity today?” *Working Paper*.
Figlio, David, Mark Rush, and Lu Yin. 2013. “Is it live or is it internet? Experimental estimates of the effects of online instruction on student learning.” *Journal of Labor Economics*, 31(4): 763–784.
Gerlitz, Jean-Yves, and Jürgen Schupp. 2005. “Zur Erhebung der Big-Five-basierten persoenlichkeitsmerkmale im SOEP.” *DIW Research Notes*, 4(10.1016).
Goodman, Joshua, Julia Melkers, and Amanda Pallais. 2019. “Can online delivery increase access to education?." *Journal of Labor Economics*, *37*(1), 1-34.
Herbert, Jennifer, and Deborah Stipek. 2005. "The emergence of gender differences in children's perceptions of their academic competence." *Journal of applied developmental Psychology*, 26(3): 276-295.
Humlum, Anders, and Emilie Vestergaard. 2024. “The Adoption of ChatGPT.” *Working Paper*.
ITU. 2019. “Measuring Digital Development: Facts and Figures 2019.”
Jack, Rebecca, and Emily Oster. 2023. “COVID-19, school closures, and outcomes.” *Journal of Economic Perspectives*, 37(4): 51–70.
Jackson, Linda A, Alexander von Eye, Hiram E Fitzgerald, Yong Zhao, and Edward A Witt. 2010. “Self-concept, self-esteem, gender, race and information technology use.” *Computers in Human Behavior*, 26(3): 323–328.
Joiner, Richard, Caroline Stewart, and Chelsey Beaney. 2015. “Gender digital divide: Does it exist and what are the explanations?” *The Wiley Handbook of Psychology, Technology, and Society*, 74–88.
Joyce, Ted, Sean Crockett, David A Jaeger, Onur Altindag, and Stephen D O’Connell. 2015. “Does classroom time matter?” *Economics of Education Review*, 46: 64–77.
Kim, Young K, and Linda J Sax. 2009. “Student–faculty interaction in research universities: Differences by student gender, race, social class, and first-generation status.” *Research in Higher Education*, 50: 437– 459.
Klein, Susan S, Patricia E Ortman, Patricia Campbell, Selma Greenberg, Sandra Hollingsworth, Judith Jacobs, Beatrice Kachuck, Averil McClelland, Diane Pollard, David Sadker, et al. 1994. “Continuing the journey toward gender equity.” *Educational Researcher*, 23(8): 13–21.
Kofoed, Michael S, Lucas Gebhart, Dallas Gilmore, and Ryan Moschitto. 2024. “Zooming to Class? Experimental evidence on college students’ online learning during Covid-19.” *American Economic Review: Insights*, 6(3): 324–340.
Li, Nai, and Gill Kirkup. 2007. “Gender and cultural differences in Internet use: A study of China and the UK.” *Computers & Education*, 48(2): 301–317.

Lichand, Guilherme, Carlos Alberto Doria, Onicio Leal-Neto, and João Paulo Cossi Fernandes. 2022. "The impacts of remote learning in secondary education during the pandemic in Brazil." *Nature Human Behaviour*, 6(8): 1079-1086.

Litman, Leib, Jonathan Robinson, Zohn Rosen, Cheskie Rosenzweig, Joshua Waxman, and Lisa M Bates. 2020. "The persistence of pay inequality: The gender pay gap in an anonymous online labor market." *PloS One*, 15(2): e0229383.

Maton, Karl. 2009. "Cumulative and segmented learning: Exploring the role of curriculum structures in knowledge-building." *British Journal of Sociology of Education*, 30(1): 43–57.

Musaddiq, Tareena, Kevin Stange, Andrew Bacher-Hicks, and Joshua Goodman. 2022. "The pandemic's effect on demand for public schools, homeschooling, and private schools." *Journal of Public Economics*, 212, 104710.

Romano, Joseph P, and Michael Wolf. 2005. "Stepwise multiple testing as formalized data snooping." *Econometrica*, 73(4): 1237–1282.

Romano, Joseph P, and Michael Wolf. 2016. "Efficient computation of adjusted p-values for resampling- based stepdown multiple testing." *Statistics & Probability Letters*, 113: 38–40.

Seel, Norbert M. 2011. *Encyclopedia of the Sciences of Learning.* Springer Science & Business Media.

Tsai, Meng-Jung, and Chin-Chung Tsai. 2010. "Junior high school students' Internet usage and self- efficacy: A re-examination of the gender gap." *Computers & Education*, 54(4): 1182–1192.

Venable, Melissa A., Eden Glass, Reece Johnson, Rebecca Long, Andrew Rice, and Liz Zvereva. 2024. "Online Education Trends Report: An Annual Survey of Students and School Administrators." *BestColleges.* https://res.cloudinary.com/highereducation/images/v1647967676/BestColleges.com/2022_Online-Education_Trends-Report/2022_Online-Education_Trends-Report.pdf

Weiser, Eric B. 2000. "Gender differences in Internet use patterns and Internet application preferences: A two-sample comparison." *Cyberpsychology and Behavior*, 3(2): 167–178.

Wong, Wan Hoong, and Elaine Chapman. 2023. "Student satisfaction and interaction in higher education." *Higher Education*, 85(5): 957–978.

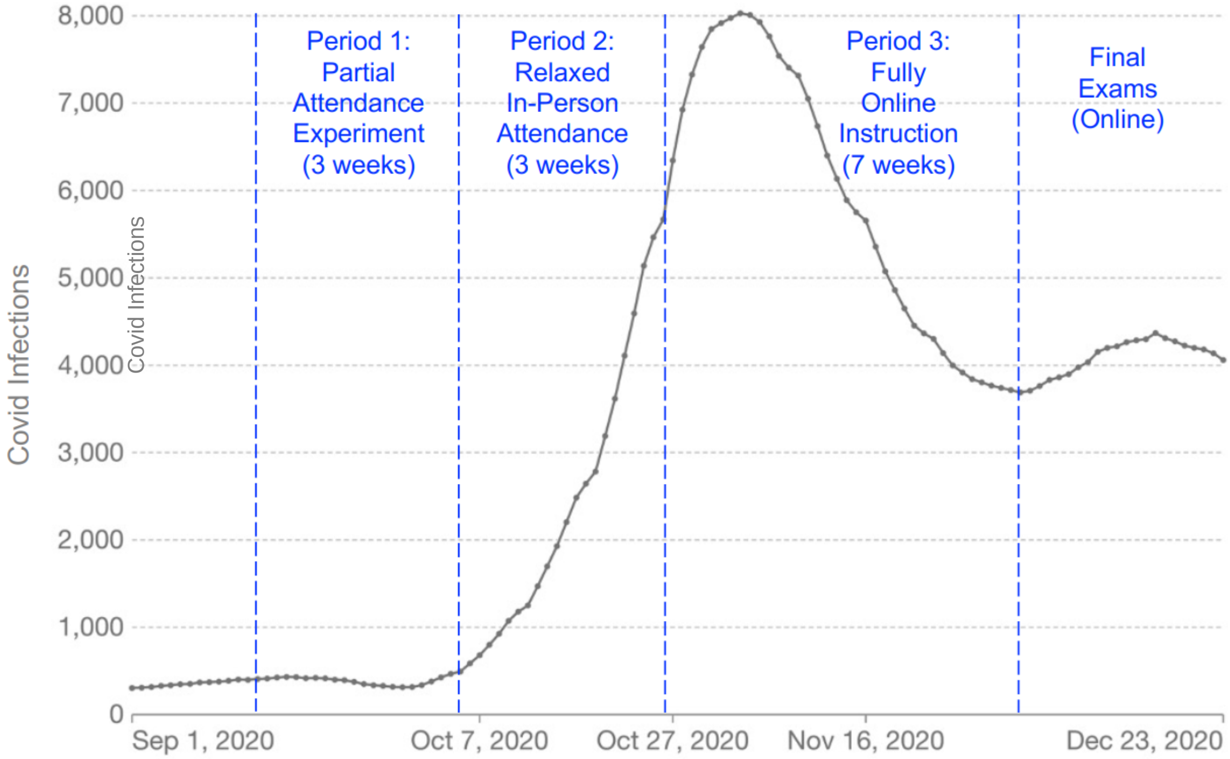
Xu, Di, and Shanna Smith Jaggars. 2013. "The impact of online learning on students' course outcomes: Evidence from a large community and technical college system." *Economics of Education Review*, 37: 46– 57.

Zhu, Chang. 2012. "Student satisfaction, performance, and knowledge construction in online collaborative learning." *Journal of Educational Technology & Society*, 15(1): 127–136.

Online Appendix

A Additional Figures and Tables

Figure A.1. A Timeline of the Experiment and the Covid Situation



Notes: The figure shows a timeline of the experiment during the fall semester of the 2020/21 academic year. The gray curve in the background shows the number of Covid infections in the country. The blue texts indicate the three periods of the semester. (1) The first three weeks were the period when we conducted the partial attendance experiment—20% of students were permitted to show up on campus at any given time. (2) During the next three weeks, the university relaxed in-person attendance restrictions. (3) During the last seven weeks of the semester, all instruction was moved online due to the surge in Covid infections. All final exams were administered online.

Table A.1. Descriptive Statistics by Student Gender

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Female Students			Male Students			Gender Difference
	N	Mean	SD	N	Mean	SD	*p*-value
Panel A: Student Characteristics (from admin data)							
Age	548	20.95	2.400	796	21.59	2.619	<0.001
International Student	548	0.162	0.369	796	0.153	0.360	0.651
Commute Time to University (Minutes)	548	40.97	38.05	796	40.27	36.41	0.735
Business/Economics/Finance Majors	548	0.518	0.500	796	0.606	0.489	0.001
Informatics Majors	548	0.097	0.296	796	0.245	0.430	<0.001
Other Majors	548	0.385	0.487	796	0.149	0.357	<0.001
Panel B: Student Characteristics (from baseline survey)							
Conscientiousness	388	4.926	0.988	504	4.696	0.917	<0.001
Extraversion	388	4.855	1.288	504	4.663	1.203	0.023
High School Grade	388	4.714	0.510	504	4.623	0.510	0.008
High School Self-Study Intensity	387	0.767	0.183	493	0.772	0.215	0.715
Group Study Preference	387	4.034	1.277	504	4.696	0.917	0.076
Panel C: Performance							
Total Exam Score	1,502	32.99	18.89	2,268	35.00	18.84	0.001
Final Course Grade	1,502	4.254	0.827	2,268	4.410	0.839	<0.001
Course Passing	1,502	0.722	0.448	2,268	0.774	0.418	<0.001
Question-Level Points	57,925	1.065	1.355	87,102	1.181	1.363	<0.001

Notes: The table shows the summary statistics of student-level characteristics and exam performance outcomes separately for the sample of women (in Columns 1–3) and men (in Columns 4–6). Column (7) shows the *p*-values derived from tests of gender difference. See Table 1 for summary statistics for the whole sample and details of student characteristics.

Table A.2. Balance Tests by Gender

Dependent Variable:	(1)	(2)
	Effect of Online Proportion	
	Women	Men
Age	0.290	-0.501***
	(0.201)	(0.179)
International Student	0.007	0.000
	(0.033)	(0.025)
Business/Economics/Finance Majors	0.069	-0.036
	(0.053)	(0.037)
Informatics Majors	-0.006	0.013
	(0.020)	(0.029)
Other Majors	-0.063	0.024
	(0.056)	(0.033)
Commute Time to University (Minutes)	3.129	-0.099
	(3.079)	(2.123)
High School Grade	-0.023	-0.019
	(0.047)	(0.035)
High School Self-Study Intensity	-0.012	0.001
	(0.016)	(0.015)
Group Study Preference	-0.136	0.001
	(0.123)	(0.088)
Conscientiousness	-0.046	0.017
	(0.098)	(0.064)
Extraversion	-0.043	-0.081
	(0.124)	(0.084)

Notes: The table shows the estimated effects of the proportion of online lectures (Online Proportion) on different baseline characteristics, separately for women and men in our sample. Each point estimate is derived from one OLS regression, in which we control for course fixed effects and cluster standard errors at the student level. We report the estimated coefficients and standard errors (in parentheses) of each baseline characteristic. **p* <.1, ***p* <.05, ****p* <.01.

Table A.3. The Impact of Online Instruction on Final Grade and Course Passing

	(1)	(2)	(3)	(4)	(5)	(6)
	Standardized Final Grade			Passing the Course		
	All	Women	Men	All	Women	Men
% Online Lectures	-0.471***	-0.473***	0.085	-0.204***	-0.204***	0.022
	(0.116)	(0.115)	(0.092)	(0.070)	(0.069)	(0.051)
% Online Lectures × Male	0.558***			0.228***		
	(0.148)			(0.086)		
Observations	3,488	1,401	2,087	3,488	1,401	2,087
R-squared	0.743	0.725	0.754	0.561	0.548	0.57
Student Fixed Effects	✓	✓	✓	✓	✓	✓
Question Fixed Effects	✓	✓	✓	✓	✓	✓

Notes: The table shows the estimated impact of online instruction on final course grade (standardized across the whole sample) and the likelihood of passing a course (i.e., at least obtaining a grade of 4). Each column represents one regression. Results are estimated with OLS regressions in Columns (1)–(3) and linear probability models in Columns (4)–(6). All regressions control for student fixed effects and course fixed effects. Standard errors in parentheses are clustered at the student level. **p* <.1, ***p* <.05, ****p* <.01.

Table A.4. Heterogeneous Impact with Multiple Testing Corrections

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Dependent Variable: Standardized Exam Score								
% Online	-0.379***	-0.174***	-0.032	-0.162**	-0.108	-0.066	-0.193**	-0.098	-0.129
	(0.085)	(0.058)	(0.116)	(0.072)	(0.087)	(0.084)	(0.082)	(0.085)	(0.091)
	[0.0001]	[0.001]	[0.957]	[0.021]	[0.521]	[0.841]	[0.013]	[0.596]	[0.347]
% Online × Male	0.409***								
	(0.108)								
	[0.0001]								
% Online × International student		0.256*							
		(0.145)							
		[0.131]							
% Online × Bus./Fin./Econ. majors			-0.132						
			(0.130)						
			[0.644]						
% Online × Higher commute time				-0.064					
				(0.106)					
				[0.918]					
% Online × Higher high school grade					-0.038				
					(0.117)				
					[0.957]				
% Online × Higher conscientiousness						-0.123			
						(0.117)			
						[0.640]			
% Online × Higher extraversion							0.127		
							(0.116)		
							[0.640]		
% Online × Higher self-study intensity								-0.051	
								(0.118)	
								[0.957]	
% Online × Higher group study preference									0.001
									(0.119)
									[0.989]
Observations	3,488	3,488	3,488	3,488	2,807	2,807	2,807	2,770	2,799
R-squared	0.839	0.838	0.838	0.838	0.840	0.840	0.840	0.840	0.840

Notes: The table shows the effects of online instruction ("% Online") interacted with different student attributes (see Table 1 for details of these characteristics). "Bus./Fin./Econ." is a binary indicator for students with business, finance, or economics majors. "Higher" means that the value of a characteristic is above the median level. Each column represents one OLS regression in which we control for both student fixed effects and course fixed effects. Standard errors are clustered at the student level and in parentheses. *p <.1, **p <.05, ***p <.01. We also present Romano-Wolf corrected p-values in brackets based on 1,000 resampling of student clusters (each student has multiple course observations) within the strata of assigned attendance groups.

Table A.5. Main Results Estimated with the Long-Term Analysis Sample

	(1) Standardized Exam Score	(2) Standardized Final Grade	(3) Standardized Exam Score	(4) Standardized Final Grade
% Online Lectures × Male	0.042	0.095		
	(0.071)	(0.095)		
% Online Lectures × Female	-0.425***	-0.523***		
	(0.093)	(0.126)		
% Online in Microeconomics × Female			-0.731***	-0.725***
			(0.231)	(0.248)
% Online in Microeconomics × Male			-0.271	-0.101
			(0.219)	(0.224)
% Online in Mathematics × Female			-0.411***	-0.639***
			(0.148)	(0.197)
% Online in Mathematics × Male			0.096	0.044
			(0.133)	(0.174)
% Online in Financial Accounting × Female			-0.416***	-0.529***
			(0.145)	(0.194)
% Online in Financial Accounting × Male			0.078	0.076
			(0.139)	(0.183)
% Online in Business Administration × Female			-0.193	-0.222
			(0.137)	(0.189)
% Online in Business Administration × Male			0.193	0.300*
			(0.123)	(0.172)
Observations	2,946	2,946	2,946	2,946
R-squared	0.831	0.742	0.831	0.742
Gender Gap (*p*-value)	0.0001	0.0001		
Gender Gap in Microeconomics (*p*-value)			0.0004	0.0002
Gender Gap in Mathematics (*p*-value)			<0.0001	0.0001
Gender Gap in Financial Accounting (*p*-value)			0.0001	0.0003
Gender Gap in Business Administration (*p*-value)			0.0013	0.0013

Notes: The table shows the estimated effects of online instruction on students' obtained exam score and final grade in different courses. The observations include the sample of 836 students (used for long-term analyses) enrolled in various courses. Each column represents one OLS regression, in which we control for student fixed effects and course fixed effects. The dependent variables are standardized across the sample to have a mean of 0 and SD of 1. Standard errors in parentheses are clustered at the student level. **p* <.1, ***p* <.05, ****p* <.01.

Table A.6. Impact on Performance for Exam Questions Covered in Different Periods of the Semester

	(1) *All*	(2) *T1*	(3) *T2*	(4) *T3*
Panel A: Obtained Points for the Question				
% Online × Female	-0.233***	-0.061	-0.196***	-0.308***
	(0.040)	(0.123)	(0.063)	(0.055)
% Online × Male	0.038	0.112	0.011	0.025
	(0.032)	(0.098)	(0.051)	(0.044)
Observations	145,027	20,715	49,835	74,477
R-squared	0.362	0.429	0.314	0.388
Student Fixed Effects	✓	✓	✓	✓
Question Fixed Effects	✓	✓	✓	✓
Gender Gap	<0.0001	0.2706	0.0107	<0.0001
Panel B: Question Correctly Answered				
% Online × Female	-0.057***	-0.007	-0.066***	-0.068***
	(0.015)	(0.038)	(0.025)	(0.022)
% Online × Male	0.003	0.014	0.000	-0.003
	(0.012)	(0.031)	(0.021)	(0.018)
Observations	145,027	20,715	49,835	74,477
R-squared	0.265	0.371	0.287	0.262
Student Fixed Effects	✓	✓	✓	✓
Question Fixed Effects	✓	✓	✓	✓
Gender Gap	0.0025	0.6665	0.0439	0.0225

Notes: The table presents the estimated effects of online instruction on the obtained points (ranging from −1.5 to 4) for various exam questions and the likelihood of correctly answering the questions. T_1 indicates questions covered in weeks 1 to 3 of the semester; T_2 indicates questions covered in weeks 4 to 6; T_3 indicates questions in weeks 7 to 13. Results in each column is derived from one OLS regression or a linear probability model, using student-question-level observations. All specifications control for student fixed effects and question fixed effects and cluster standard errors (in parentheses) at the student-by-question level. *p <.1, **p <.05, ***p <.01.

Table A.9. Exposure to Online Instruction and Endline Survey Response

	(1)	(2)	(3)	(4)
	Dependent Variable: Endline Survey Response			
Panel A: All Students				
% Online in Financial Accounting	0.030			
	(0.080)			
% Online in Business Administration		-0.105		
		(0.097)		
% Online in Mathematics			0.029	
			(0.076)	
% Online in Microeconomics				-0.062
				(0.084)
Female	0.153***	0.117***	0.106***	0.111***
	(0.038)	(0.034)	(0.035)	(0.033)
Observations	820	999	915	1,036
R-squared	0.029	0.019	0.027	0.029
Panel B: Female Students				
% Online in Financial Accounting	0.010			
	(0.110)			
% Online in Business Administration		-0.210		
		(0.153)		
% Online in Mathematics			-0.038	
			(0.120)	
% Online in Microeconomics				-0.105
				(0.113)
Observations	336	410	328	428
R-squared	0.006	0.017	0.039	0.035
Panel C: Male Students				
% Online in Financial Accounting	0.052			
	(0.122)			
% Online in Business Administration		-0.013		
		(0.127)		
% Online in Mathematics			0.078	
			(0.100)	
% Online in Microeconomics				-0.035
				(0.128)
Observations	484	589	587	608
R-squared	0.021	0.015	0.021	0.019

Notes: The table presents the estimated impact of the proportion of online lectures in each course (if enrolled in the course) on the response rate for the endline survey. The dependent variable is a binary indicator for completing the endline survey. Panel A includes all students; Panel B/C includes female/male students in our sample. Each column in each panel represents one linear probability model that controls for major, nationality, age, commute time, and an indicator for whether being in the first year. Robust standard errors are in parenthesis. *p <.1, **p <.05, ***p <.01.

Table A.10. The Impact of Online Instruction on Performance by Course

	(1)	(2)	(3)	(4)
	Standardized Score		Standardized Grade	
Panel A: Female Students				
% Online Lectures in Microeconomics	-0.920***	-0.992***	-0.772*	-0.747**
	(0.355)	(0.306)	(0.411)	(0.355)
% Online Lectures in Mathematics	-0.408**	-0.334*	-0.961***	-0.829***
	(0.174)	(0.194)	(0.330)	(0.276)
% Online Lectures in Financial Accounting	-0.007	-0.094	-0.040	-0.217
	(0.215)	(0.177)	(0.279)	(0.258)
% Online Lectures in Business Administration	-0.157	-0.152	-0.119	-0.140
	(0.124)	(0.162)	(0.254)	(0.245)
Observations	1,502	1,401	1,502	1,401
R-squared	0.575	0.843	0.131	0.726
Panel B: Male Students				
% Online Lectures in Microeconomics	-0.033	-0.030	0.072	0.087
	(0.305)	(0.260)	(0.335)	(0.258)
% Online Lectures in Mathematics	0.073	0.006	0.162	0.071
	(0.128)	(0.156)	(0.230)	(0.203)
% Online Lectures in Financial Accounting	0.154	-0.131	0.226	-0.173
	(0.194)	(0.152)	(0.247)	(0.195)
% Online Lectures in Business Administration	-0.016	0.207	-0.064	0.273
	(0.104)	(0.138)	(0.210)	(0.187)
Observations	2,268	2,087	2,268	2,087
R-squared	0.550	0.837	0.157	0.754
Course Fixed Effects	Yes	Yes	Yes	Yes
Student Controls	Yes	–	Yes	–
Student Fixed Effects	No	Yes	No	Yes

Notes: The table shows the estimated impact of online instruction on students' performance in four courses. The outcome variable is the standardized exam score in Columns (1)–(2) and the standardized course grade in Columns (3)–(4). Each column represents one OLS regression. The specifications in Columns (1) and (3) control for course fixed effects and student characteristics including age, international status, log commute time, and first-year indicator. The specifications in Columns (2) and (4) control for both course fixed effects and student fixed effects. Standard errors in parentheses are clustered at the student level. *p <.1, **p <.05, ***p <.01.

Table A.11. Question-Level Teaching Mode and Performance

	(1)	(2)	(3)	(4)
	Obtained Points		Answered Correctly	
Online-Only Lecture	0.009		0.002	
	(0.023)		(0.007)	
Online-Only Lecture × Male		0.002		-0.005
		(0.029)		(0.009)
Online-Only Lecture × Female		0.021		0.012
		(0.037)		(0.011)
Observations	20,715	20,715	20,715	20,715
R-squared	0.429	0.429	0.371	0.371
Gender Gap (*p*-value)		0.671		0.239

Notes: The table examines whether students perform differently for questions covered in an online-only lecture and questions covered in an in-person lecture, within the first weeks of the semester. Each column is estimated with one OLS regression or one linear probability model, using student-by-question level observations. The dependent variable is the raw points obtained for a question in Columns (1) and (2) and an indicator for whether a question is correctly answered in Columns (3) and (4). In all regressions, we control for student fixed effects and question fixed effects and cluster standard errors (in parentheses) at the student-by-question level. None of the estimates are statistically significant.

Table A.12: Summary of Prior Studies

Authors (Year) and Journal	**University**	**Course(s)**	**Treatment vs. Control**	**Empirical Strategy**	**% of Women**	**Sample size**
Alegria et al. (2023). *Education Economics*	Private university in Spain	*Applied business math & data analysis for economics*	Online vs. in-person	IV	41%	403
Coates et al. (2004). *Economics of Education Review*	Three universities in the US	*Principles of economics*	Online vs. in-person	IV	48%	126
Alpert et al. (2016). *American Economic Association, Papers & Proceedings*	Public university in the US	*Principles of microeconomics*	Online vs. hybrid vs. in-person	RCT	44%	323
Bettinger et al. (2017). *American Economic Review*	For-profit university in the US	Many courses	Online vs. in-person	IV	47%	2,323,023
Xu and Jaggars (2013). *Economics of Education Review*	34 community colleges in the US	Many courses	Online vs. in-person	IV	53%	116,830
Kofoed et al. (2024). *AER: Insights*	Selective Military Academy in the US	*Introductory economics*	Online vs. in-person	RCT	23%	551
Figlio et al. (2013). *Journal of Labor Economics*	Selective university in the US	*Principles of microeconomics*	Online vs. in-person	RCT	55%	296
Joyce et al. (2015). *Economics of Education Review*	Public university in the US	*Introductory microeconomics*	Less vs. more in-person; online materials always available	RCT	45%	656
Cacault et al. (2021). *Journal of the European Economic Association*	Public university in Switzerland	Several courses (economics, math, management, HR)	No vs. partial vs. unlimited **access to streaming**; in-person attendance always available	RCT	51%	23,766
Bowen et al. (2014). *Journal of Policy Analysis and Management*	Six public universities in the US	*Statistics*	Hybrid vs. in-person	RCT	58%	431
Shan et al. (2025). ***Working Paper***	Public university in Switzerland	Several courses (economics, math, business, finance)	Online vs. in-person	RCT	41%	3,488

Appendix B. Deviations from the Pre-Registration and Pre-Analysis Plan

Our experiment was registered on the AEA RCT Registry (ID: AEARCTR-0006538). In the preregistration, we listed four questions (related to different outcomes) to be answered with this experiment:

- Question 1: Do students perform better on exam questions covered in an online lecture versus an in-person lecture?
- Question 2: Does online instruction affect attendance, course dropout, course passing, and grades?
- Question 3: Does online instruction affect students' social interactions, satisfaction with course, study habits, and learning experience?
- Question 4: Does online instruction affect study dropout, completion of the first year, elective course choices, and major switching?

In the paper we have analyzed all four questions:

(1) Regarding Question 1, we show how online instruction affects question-level performance in Section 5.3.

(2) Regarding Question 2, we use course-level performance as the primary outcome and show the effects of online instruction mainly in Section 5.1. We also examine the effects on attendance in Section 6.1 and Section 6.2.

(3) Regarding Question 3, we analyze how online instruction affects different dimensions of learning experiences in Section 6.3.

(4) Regarding Question 4, we show how online instruction affects longer-run performance and study dropout in Section 5.5. However, this paper presents results with slightly different focuses and omits some secondary outcomes from the analysis due to data limitations that were unaware of at the time of preregistration. Below, we discuss and justify these deviations in greater detail.

Deviation (1): Initially, Question 1 represented our primary empirical strategy. It relies on the student-by-question level variation in the model of instruction to examine how online instruction affects question-level performance. However, due to unforeseen changes in the Covid-19 situation, our experiment was only strictly implemented for the first three initial weeks of the semester. This period covers only 13% of all exam questions, and questions covered later in the semester do not vary in the mode of instruction. This ruled out to use of student-by-question level variation as the primary strategy. As discussed in Section 5.3, we nevertheless show the results in Appendix Table A.11.

Deviation (2): Out of all the course-level outcomes listed for Question 2, we do not explicitly show results on course dropout because (1) we do not directly observe it in administrative data—what we do observe is students' decision to attend or skip the final exams which were scheduled online; and (2) according to this measure, the short-run dropout rate is relatively low (about 8%), which limits our power to detect any effect. When testing how online instruction assigned at the start of semester affects course dropout, we find very small and statistically insignificant effects for both men and women. Specifically, using the specification that controls for student fixed effects and course fixed effects, we find that increasing online proportion by 10 percentage points lowers women's dropout rate by 0.08 percentage points (p-value = 0.595) and increases men's dropout rate by 0.02 percentage points (p-value = 0.872). We therefore decided to focus our analysis on students who have attended the final exams of enrolled courses—comprising 92% of full sample.

Deviation (3): Although we pre-registered various long-term outcomes, we decided to focus on study dropout, because it is the most accurate outcome we observe in our acquired administrative data. There is no clear-cut way to define first-year completion and elective course selection, given that students are enrolled in different major or minor programs, and because many courses can be postponed or retaken. Since we do not have access to administrative data from outside of the business school, we do not observe detailed records of major switching.

Deviation from the pre-analysis plan: In the pre-analysis plan, we listed nine dimensions of sub-group analysis for the following student attributes: gender, previous achievement, nationality, self-study habits, group study preference, conscientiousness, extraversion, commute time, and major. We have analyzed heterogeneity by all these attributes, but only the heterogeneity by gender is highly statistically significant and robust to multiple testing corrections (as discussed in Section 5.2).