CESifo WORKING PAPERS

8671
2020

November 2020

Can Peer Mentoring Improve Online Teaching Effectiveness? An RCT during the COVID-19 Pandemic

David Hardt, Markus Nagler, Johannes Rincke

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8671*

Can Peer Mentoring Improve Online Teaching Effectiveness? An RCT during the COVID-19 Pandemic

Abstract

Online delivery of higher education has taken center stage but is fraught with issues of student self-organization. We conducted an RCT to study the effects of remote peer mentoring at a German university that switched to online teaching due to the COVID-19 pandemic. Mentors and mentees met one-on-one online and discussed topics like self-organization and study techniques. We find positive impacts on motivation, studying behavior, and exam registrations. The intervention did not shift earned credits on average, but we demonstrate strong positive effects on the most able students. In contrast to prior research, effects were more pronounced for male students.

JEL-Codes: I200, I230, J240.

Keywords: online education, EdTech, COVID-19, mentoring, higher education.

David Hardt
Friedrich Alexander University Erlangen-Nuremberg / Germany
david.hardt@fau.de

Markus Nagler
Friedrich Alexander University Erlangen-Nuremberg / Germany
markus.nagler@fau.de

Johannes Rincke
Friedrich Alexander University Erlangen-Nuremberg / Germany
johannes.rincke@fau.de

October 30, 2020
This field experiment was pre-registered at the AEA Social Science Registry under the ID AEARCTR-0005868 and has the approval from the ethics commission as well as from the data protection officer at the university where the experiment took place. Uschi Backes-Gellner, Fabian Dehos, Ulrich Glogowsky, Joshua Goodman, Philipp Lergetporer, Martin Watzinger, Kathrin Wernsdorf, and Martin West provided helpful comments and suggestions. We thank Jens Gemmel for excellent research assistance. We thank Sophie Andresen, Helen Dettmann, Eva-Maria Drasch, Daniel Geiger, Nina Graßl, Jana Hoffmann, Lukas Klostermeier, Alexander Lempp, Lennart Mayer, Jennifer Meyer, Annabell Peipp, Tobias Reiser, Sabrina Ried, Niklas Schneider, and Matthias Wiedemann for their excellent work as student mentors. Nagler gratefully acknowledges funding by the Joachim Herz Foundation through an Add-On Fellowship. Rincke gratefully acknowledges funding by the Innovationsfonds Lehre at the University of Erlangen-Nuremberg.

1 Introduction

Online delivery of tertiary education is on the rise throughout the world. Currently, the COVID-19 pandemic has forced virtually all education institutions to switch to online teaching. The literature on online teaching has generally found this format of teaching to be somewhat inferior to classical classroom-based teaching (Brown and Liedholm, 2002; Figlio et al., 2013; Joyce et al., 2015; Alpert et al., 2016; Bettinger et al., 2017b). This may be due to problems of disorganization among students in online teaching and a lack of personalization, as has been argued for massive open online courses (so-called MOOCs; see e.g. Banerjee and Duflo, 2014; McPherson and Bacow, 2015). Switching to online teaching may thus aggravate a situation in tertiary education in which many students struggle to successfully complete their studies in time (Himmler et al., 2019).[1] Accordingly, students expect and experience negative consequences of the COVID-19-induced shift to online teaching for their study outcomes (Aucejo et al., 2020; Bird et al., 2020).

One way to improve on outcomes of online education could be to assist students in their self-organization and enhance personalization by providing online peer-to-peer mentoring. Leading universities have launched student coaching programs to support their students during the COVID-19 crisis, and the American Economic Association (AEA) recommended in the early phase of the pandemic that graduate programs should set up "more rigorous mentoring systems for students who will not be able to benefit from the usual sorts of interactions with peers and professors".[2] However, evidence on the effectiveness of such mentoring programs is generally scarce for higher education and nonexistent for online teaching environments, where they may be particularly helpful.

In this paper, we report the results of a randomized controlled trial to study the effects of peer mentoring at a large German public university that, due to the COVID-19 pandemic, switched to online teaching for the summer term 2020. The context of the trial was the core undergraduate program at the university's School of Business and Economics. Our sample comprises of 691 students enrolled in the second term of the program. To assess the effectiveness of the peer mentoring program, we combine

[1] A large share of students never obtain a degree, and those who do often take much longer than the design of the program would suggest. For instance, data from the National Center for Education Statistics show that in the United States, less than 40 percent of a cohort entering four-year institutions obtain a bachelor's degree within four years. Data on other countries document that similar problems are widespread. Overall, in OECD countries the completion rate at the tertiary level is only 70 percent. See, e.g., http://nces.ed.gov/programs/digest/d13/tables/dt13_326.10.asp.

[2] An example for such programs is the MIT's Student Success Coaching program, see http://news.mit.edu/2020/student-coaching-calls-pandemic-0501 for details. The AEA published the recommendations on mentoring (together with other guidelines on graduate programs) on May 11, 2020, via email to all members of the association.

administrative student data with complementary survey data that we collected before the start of the examination period in the summer term.

We designed a structured peer mentoring program that focused on providing students with general study skills, such as self-organization in a learning-from-home environment, weekly study schedules, and techniques on how to study effectively. Mentors and mentees met one-on-one online. Overall, we designed a mentoring program consisting of five structured meetings, which took place every two to three weeks. In every meeting, mentors would discuss specific topics, such as mentees' weekly study schedules, using materials and templates provided by us. The meetings would then also involve follow-up discussions on how students were coping with putting previous inputs to practice. They would also involve discussions on topics that mentees would suggest. Importantly, we specifically instructed mentors not to discuss any coursework or any specific content-based problems with mentees. As mentors, we hired 15 students from a more advanced term in the same study program as the mentees as student assistants. Importantly, the type of mentoring we offered could be scaled up easily and at low cost: Including one additional mentee into the program for a three-month teaching period would cost just about €60.[3]

The setting in which the mentoring program took place is typical for public universities across the developed world. In each winter semester, about 700 active students enroll in the three-year (six-semester) bachelor's program *Economics and Business Studies*.[4] In each of the first two semesters, students are supposed to pass exams in six core courses, each of them worth five credits. Administrative data from the academic year 2018/19 shows that even in regular times, many students underperform relative to the suggested curriculum in the first study year: after the first semester, only 59 percent of students still enrolled at this point in time have completed courses worth at least 30 credits. The curriculum for the second semester comprises some courses involving more rigorous methods relative to the first semester. As a result, the students' performance typically further decreases in the second semester: in 2019 (the year prior to the intervention), only about 25 percent of students had completed 60 credits at the end of their first study year. A key advantage of our setting is that the summer term 2020 was conducted entirely online at the university because the German academic year starts and ends later than is common internationally. Thus, all classes offered to students in the summer term were online-only classes with zero in-person teaching.

[3]All mentors were employed for three months during the summer term, based on work contracts with for hours per week and a monthly net pay of about €160. Employer wage costs were about €200 per month and mentor.

[4]The program is broad and can lead to specializations in business administration, economics, information systems, and business and economics education.

We find that the peer mentoring program improved students' outcomes and study behavior. Students in the treatment group register for significantly more exams. They sit and pass some of these, such that the average effects on passed credits are attenuated. Students' GPA is unaffected by our intervention. Our baseline intent-to-treat effect shows an effect of around 1.4 additional credits for which students register, which is equivalent to 28% of an additional course. Given that the take-up rate in the treatment group is around 40%, the associated instrumental variable estimates show an increase of 3.4 credits, around 70% of an additional course. Of these, they pass 0.5 on average (IV estimation: 1.3 more credits passed). In the survey, students in the treatment group report significantly higher overall motivation to study, are more likely to report having studied continuously throughout the term, and are more likely to think they provided enough effort during the summer term to reach their goals. The largest effects are on students' motivation, where instrumental variables estimates show increases of around 0.5 points on a five-point Likert scale, around 17% relative to the average. In contrast, students' views on departmental services or on online teaching in the summer term or in general seem unaffected.

The heterogeneity of our effects is in contrast to prior research on the impacts of peer mentoring in higher education. First, while prior research suggests that weaker students struggle most in online learning environments (e.g., Figlio et al., 2013; Bettinger et al., 2017b), our results show that good students benefit more from the mentoring program. For example, while the bottom two terciles of the distribution of credits earned in the winter term are unaffected, the upper tercile passes more than one additional course in the summer term with treatment than without. In sum, for students in the upper tercile, the probability of reaching the designated performance goal of 60 credits after two terms increases by around 9 percentage points, or 18% relative to the mean. This is interesting since many evaluations of in-person mentoring programs in higher education that find no effect of academic support services explicitly exclude good students (e.g., Angrist et al., 2009). Second, in line with the evidence from prior research, we find that female students are more likely to sign up for the mentoring program conditional on invitation. However, in contrast to prior research (e.g., Angrist et al., 2009), female students are not more likely to benefit from the intervention: if anything, male students show stronger treatment effects of the program.

Our paper contributes to the growing but small literature on the online education production function. This literature has generally found online teaching to be less effective than classroom-based teaching (see, e.g., Brown and Liedholm, 2002; Figlio et al., 2013; Joyce et al., 2015; Alpert et al., 2016). A key paper in this literature is Bettinger et al. (2017b), who analyze students at a for-profit university that offers courses both in-person and online. Using distance to this university interacted with online course availability as an instrumental variable for online course taking, they find that students perform worse in online courses and that student achievement becomes more variable in online environments. This may be due to problems of disorganization among students in online teaching, as has been argued for massive open online courses (e.g. Banerjee and Duflo, 2014; McPherson and Bacow, 2015). The literature on specific aspects of the online education production function is however small. Bettinger et al. (2017a) study students at a for-profit college and randomly allocate students to differentially sized online classrooms and find little effects. We contribute to this literature by providing the first evidence on the effectiveness of peer mentoring programs for online higher education, to our knowledge.[5]

Our paper also contributes to the literature on the effectiveness of mentoring interventions in higher education. The literature on mentoring has so far mostly focused on settings before the onset of tertiary education (see, e.g., Lavy and Schlosser, 2005; Rodriguez-Planas, 2012; Oreopoulos et al., 2017).[6] In tertiary (classroom-based) education, the results of mentoring interventions seem promising, although the literature is not large. Bettinger and Baker (2014) show that a student coaching service focusing on aligning long-term goals and self-organization and providing study skills increased university retention.[7] In a paper that is close in spirit to ours, Angrist et al. (2009) test the impact of a combination of academic support services and financial incentives on students' GPA. They find that this raises performance among female students. We replicate their finding that female students are more likely to sign up for mentoring services, but do not find a similar heterogeneity in outcomes. If anything, men seem to benefit more from our mentoring program. Also in contrast to our results, they find no effects for academic support services that are not combined with financial incentives. In comparison to their intervention, our program is targeted more towards individual mentor-mentee interactions, is more structured and specific regarding the advice given to mentees, and our intervention takes place in an online environment where, arguably, mentoring services may be more important. We thus contribute to this literature by providing the first evidence on the effectiveness of (peer) mentoring in an online context, to the best of our knowledge, and by extending the small experimental literature on the effects of mentoring in higher education.

[5]For a paper on the potential extensive margin effects of online education, see Goodman et al. (2019). For research on the value of an online degree, see Deming et al. (2016). For research on potential cost and competitive effects of online higher education on offline higher education, see Deming et al. (2015) and Deming et al. (2018), respectively.

[6]There is an additional related literature on assistance provision in higher education (see, e.g. Bettinger et al., 2012). For research on mentoring in other settings, see, e.g., Lyle and Smith (2014).

[7]Castleman and Page (2015) provide evidence of an effective text messaging mentoring for high-school graduates.

Finally, our paper contributes to the literature on effective education responses to the COVID-19 pandemic. Most research in this area has focused on primary or secondary education (e.g., Bacher-Hicks et al., 2020; Grewenig et al., 2020). The paper most closely connected here is Carlana and Ferrara (2020), who conducted an RCT assigning middle school students in Italy an online mentor during the pandemic. In line with our results, they find that online tutoring improves student performance and student well-being. We contribute to this research area by providing evidence on the effectiveness of an online mentoring program in higher education in a setting where teaching went online due to COVID-19. Although virtually all universities in the world have switched to online teaching due to the pandemic, evidence on useful measures to improve the effectiveness of online teaching in this context remains nonexistent. This is despite early evidence suggesting that the shift led to worse outcomes in higher education (Bird et al., 2020).

The remainder of this paper is structured as follows. In the next section, we provide details on the setting and the experimental design. Section 3 informs about the data and the empirical setup. We discuss the treatment effects on administrative student outcomes in Section 4. This includes average effects as well as heterogeneous effects by prior student performance and gender. Section 5 shows treatment effects on survey outcomes, including motivation and study behavior. The final section concludes.

2 Experimental Setting and Design

2.1 Experimental Setting

The setting of our experimental study is typical of public universities in Europe and the Western world. The undergraduate study program *Economics and Business Studies* at the School of Business and Economics at the university where the trial was implemented requires students to collect 180 credits to graduate. Students are expected to graduate after three years (six semesters). The study plan assigns courses worth 30 credits to each semester. Administrative data show that a large share of students do not complete 30 credits per semester, delaying their graduation. At the same time, survey data collected from an earlier cohort of students suggests that most students do not work full-time even if one aggregates the hours studied and the hours worked to earn income.[8] The salient study plan and target of achieving 30 credits per term, the fact that most students do register for exams worth these credits, and the fact that students do not seem to work enough to pass these exams suggests that many students

[8]On average in the first two semesters, survey participants spend about 13.3 hours per week attending courses, about 9.8 hours self-studying, and 7.5 hours to earn income.

have problems in self-organizing and/or studying efficiently. Most likely, given prior findings on such problems in online education, these issues were exacerbated by the switch to online teaching. This is where our program was supposed to intervene.

Due to the COVID-19 pandemic, in the summer term 2020 all courses of the School of Business and Economics were conducted in online format. To this end, the university acquired licenses of *Zoom*, an online video conference tool used widely in academic settings during the pandemic to enable all lecturers to give online classes and seminars. While the exact implementation of online teaching differed by subject and instructor, this makes the setting similar to the setting of other academic institutions around the globe during this pandemic.

A key advantage of our setting is that the summer term 2020 was conducted entirely online at the university because the German academic year starts and ends later than is common internationally. Thus, students did not first take classes in-person before switching to online classes as in other universities where the spring or summer term had already started when the pandemic accelerated. The setting is thus cleaner than would be possible in other settings since spillovers from in-person to online teaching can be ruled out.

2.2 The Mentoring Program

In the first week of the semester, students in the treatment group were informed via e-mail about the launch of a new mentoring program designed specifically for students in the second semester of the study program. They were invited to register for the program through a webpage.[9]

The mentoring program focused on self-organization and was supposed to make mentees aware of potential problems and pitfalls of studying online. We designed the mentoring program to involve five one-on-one online meetings between mentors and mentees. Each meeting was supposed to last between 30 and 45 minutes. The average length of meetings as reported to us by the mentors was around 40 minutes. For each of the meetings, we provided mentors with structured information on how to conduct the session.

The first meeting was meant to focus on mentees' expectations regarding their performance in the second term, and to contrast these expectations with average performance figures from previous student cohorts. The mentor was also instructed to provide practical advice on how to self-organize when working from home. In the

[9]The page asked for the students' consent to use their personal information for research purposes in anonymized form and for their consent to pass along their name and e-mail address to their mentors. We sent reminder e-mails to students in the treatment group who did not register for the program within two days.

second meeting, mentors and mentees formulated specific goals that the mentee aimed to achieve in the term. This included aims regarding study effort (time schedule for the study week, see Figures A.1 and A.2 in the Appendix) and courses to be taken. It also included performance-based goals (number of exams to pass). The third meeting was designed to focus on exam preparation (discuss timing of scheduled exams, reflect on implications for the mentee's preparation). The main topic of the fourth meeting was how to study effectively. This included the presentation of a simplified four-stage learning model (see Figure A.3 in the Appendix) and how to implement the proposed learning strategies in practice. In the fifth and final meeting, the mentor and the mentee mainly discussed the mentee's exam preparation, including a time schedule that provided the mentee with guidance on how to specifically prepare for exams. In all meetings, besides the main topics mentioned, the mentor and the mentee were instructed to discuss current general issues that the mentee was facing.[10] To limit the risk of spillovers, we instructed all mentors to make sure that the information was only provided to mentees and not to other students.

In the control group, there was no mentoring. However, the School of Business and Economics provided general information on the topics that we focus on in the mentoring for all students through its website. This included advice on how to work from home and general information on all issues regarding the online implementation of courses.

2.3 Recruitment and Training of Mentors

For administrative reasons, we had to initiate the hiring of the peer mentors about 4 weeks before the start of the program. In total, we hired 15 mentors. Work contracts were specified such that each mentor would handle a maximum of 10 mentees. The mentoring program's maximum capacity was therefore 150 students.

All mentors were students who successfully completed the first year of studies and during the summer term were enrolled in the fourth semester of the study program. Thus, all mentors were in the upper tercile of the distribution of credits earned until the summer term. They had rather good GPAs, above average high-school GPAs, and were likely to work in student jobs next to their studies. Among all applicants, we selected those that we felt were most able to work as mentors for students in less advanced cohorts. Eight of the mentors were females and seven were males.

[10]The mentors were asked to take brief notes about the content of the discussions during each meeting. We provided mentors with some structure for the notes in advance. Mentors were also instructed to prepare thoroughly for every individual meeting by recapturing the short notes they gathered during the prior meeting.

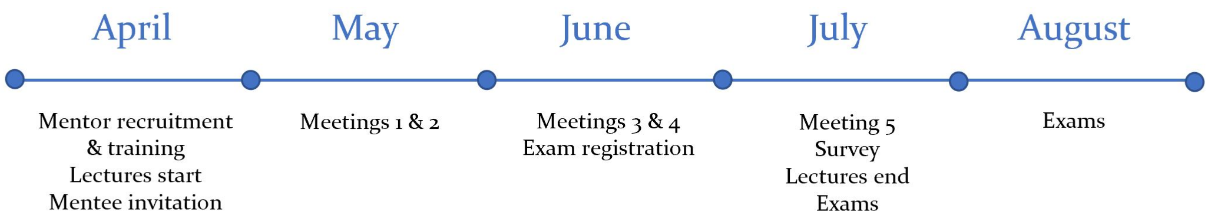


Figure 1: Timeline of Intervention

Note: This figure shows the timeline of our experiment.

Shortly before the start of the mentoring program, all mentors took part in an online kick-off meeting. In the kick-off meeting, the research team explained the purpose and the general structure of the program and laid out the planned sequence and contents of the mentoring sessions to be held with each mentee. The mentors could also ask questions. The mentors were not informed about the fact that the program was implemented in the context of an experiment. Mentors were informed about the fact that the program's capacity was limited and that a random subset of all students in the second term was invited to participate.

On the next day, all mentors took part in a training given by professional coaches. The training focused on communications skills and took about five hours (excluding breaks). Three weeks after the start of the program, the mentors took part in a short supervision meeting (about one hour) with the coaches. In addition, the members of the research team sent regular e-mails to the mentors (one e-mail before each of the five waves of meetings) and answered questions in response to individual queries by the mentors. Short feedback conversations also took place, mainly for us to get a sense on how the program was being implemented.

An overview of the timing of the project is displayed in Figure 1.

2.4 Sampling

About 850 students enrolled for the study program *Economics and Business Studies* in the winter term of 2019. We excluded from the experiment students who dropped out after the first semester, who were not formally in their second semester in the summer term 2020, for example because of having been enrolled at another university before and having already completed courses from the first or second semester of the study program without having taken these exams at the university, and who completed less than a full course (5 credits) in the first term.[11] This leaves us with 694 students entering the second term. We randomly assigned half of the students to

[11]In Germany, some students enroll at a university because as students they have access to heavily subsidized health insurance.

treatment and the other half to control. We used a stratified randomization scheme with gender and number of credits completed in the first term as strata variables. After the intervention ended, we had to drop another three students from the sample who in the meanwhile got credited for second-term courses earned elsewhere.[12] Our final sample thus consists of 691 students.

Because of the fixed capacity of the program and the (ex ante) unknown take-up rate, we first invited students sampled into treatment who did complete up to 30 credits in their first term (369 students). We then successively invited three further groups of students sampled into treatment according to the number of credits earned in the first semester, until all 344 students sampled into treatment got an invitation email. In total, 142 students from the treatment group signed up for the mentoring program.

Assignment of Students to Mentors

We randomly assigned students who signed up for the program to mentors. In order to achieve a balanced mix of mentee-mentor pairs in terms of gender, we used the mentees' gender as a strata variable in the assignment. Out of the 15 mentors, eight were females and seven were males. Among students registered for the program, about 54 percent were female. As a result, the number of mentee-mentor pairs in each of the mentee-mentor gender combinations was similar.

3 Data and Empirical Strategy

3.1 Data

Administrative Data

We collected administrative data from the university in mid October 2020 to measure all outcomes related to exam participation and academic achievement, both in the winter term 2019 and in the summer term 2020 (i.e., all of the students' first study year). Our outcomes of interest are, first, the number of credits (students receive five credits for each examination that they pass) for which students register as a measure for attempted examinations. This may be interpreted as a measure of student effort. Our primary outcome is, second, credits earned in the second term. This variable measures most directly the students' academic achievement during the term in which the intervention took place. Note however that this might be a slow-moving variable since study effort has cumulative gains over time. Following Angrist et al. (2009), we

[12] Students are free when to hand in certificates on credits earned elsewhere. As a result, such credits often show up with some delay in the administrative data.

did not exclude students who withdrew from the sample. Students who withdrew before registering for or earning any credits in the second term were coded as having zero attempted and earned credits, respectively. We do not impute a GPA for these students.

Third, we examine the impact on students' GPA for passed courses, running from 1 (passed) to 4 (best possible grade).[13] Given that we expect (and find) impacts of the treatment on the prior two main outcomes, treatment effects on GPA are not directly interpretable, though. This is in contrast to Angrist et al. (2009), whose main measure of academic achievement is students' GPA. The reason for this difference is that in the German university system, students are typically free to choose the timing of taking their courses even when a core curriculum is suggested. In addition and as outlined above, many students do not attempt to complete the core curriculum in the suggested time period, making the extensive margin decision how many courses to take more relevant than in the U.S. context.

The exams were scheduled after the end of the teaching period and took place between end of July and September 2020. All examination results had to be reported to the department administration by the end of September. In addition, the university provided us with background information on individual students. The individual characteristics include information on enrollment, gender, age, type of high school completed, and information on high-school GPA (running from 1 as the worst to 4 as the best grade).[14]

Survey Data

After the end of the intervention (i.e., after the fifth round of mentee-mentor meetings was completed), we invited all students in the experimental sample (i.e., both from the treatment and the control group) to an online survey. The survey was conducted on an existing platform at the department that is frequently used to survey students. Students who completed the survey, which lasted around ten minutes, received a payoff of €8.00. The survey elicited the students' assessment of their own study effort, their satisfaction with the department's effort to support online learning during the teaching term, views on online teaching in general, and beliefs about one's own academic achievement. The full set of survey questions is shown in Online Appendix C.1. We use all survey responses submitted until the official beginning of the examination period to avoid spillover effects from exams to survey data. Overall, 404 students (58.5% of the main sample) participated in the survey.

[13] In Germany, a reversed scale is used, with 1 being the best and 4 being the worst (passing) grade. We recoded the GPA to align with the U.S. grading system.

[14] We again recoded the GPA to align with the U.S. system.

	Control	Treatment	Difference	Std. diff.
Female	0.46	0.47	0.01	0.01
	(0.50)	(0.50)	(0.04)	
Age	21.29	21.26	-0.03	-0.01
	(2.48)	(2.69)	(0.20)	
High-school GPA	2.37	2.38	0.01	0.01
	(0.57)	(0.61)	(0.05)	
Top-tier high-school type	0.76	0.74	-0.01	-0.02
	(0.43)	(0.44)	(0.03)	
Foreign univ. entrance exam	0.07	0.08	0.02	0.04
	(0.25)	(0.27)	(0.02)	
Earned credits in first term	25.23	25.26	0.02	0.00
	(9.27)	(8.93)	(0.69)	
First enrollment	0.63	0.68	0.05	0.08
	(0.48)	(0.47)	(0.04)	
Part-time student	0.09	0.08	-0.00	-0.01
	(0.28)	(0.28)	(0.02)	
Obs.	347	344	691	691

Note: This table shows means of administrative student data (standard deviations in parentheses) by treatment status, together with differences between means and corresponding standard errors (in parentheses) and standardized differences. In the line where we report high-school GPA we need to drop 11 observations where we do not have information on students' high-school GPA.

Table 1: Summary Statistics by Treatment Status

3.2 Balancing Checks and Take-Up

Balancing

Table 1 reports differences in means (together with standard errors) and standardized differences on students' individual characteristics. The characteristics included comprise gender, age (in years), high-school GPA, a dummy for the most common type of high school certificate ("Gymnasium"), a dummy for students who obtained their high school certificate abroad, credits earned in the first term, a dummy for students who are in their first year at university, and a dummy for full-time students.[15] As can be seen from Table 1, the treatment and control groups were well balanced across all individual characteristics.

[15]Students can be in the first year of the study program, but in a more advanced year at university if they were enrolled in a different program before. About 10% of students are enrolled as part-time students because their university education is integrated into a vocational training program.

Dependent Variable:	Sign-up	Sign-up w/o dropouts		Sign-up	
	overall	before first meeting	any time	Female	Male
	(1)	(2)	(3)	(4)	(5)
Treatment	0.41***	0.37***	0.32***	0.47***	0.36***
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)
Obs.	691	691	691	324	367

Note: This table shows results of regressions of program take-up on initial treatment assignment controlling for student gender (where possible) and credits earned in the winter term. Column (1) uses initial program sign-up as the dependent variable. Column (2) uses program sign-up among those who met at least once with their mentors as the dependent variable. Column (3) uses an indicator of whether students met five times with their mentors as the dependent variable. Columns (4) and (5) use the same dependent variable as Column (1) but split the sample into female and male students, respectively. Standard errors are robust. * p<0.10, ** p<0.05, *** p<0.01

Table 2: Take-Up

Take-Up

Students who registered for the peer mentoring program could drop out at any time with no penalty. Of the 142 students who signed up for the program, 128 met at least once with their mentors. 119 students met at least three times with their mentors. Overall, 109 students met five times with their mentors. Table 2 shows the take-up of our program and the first stage of our instrumental variable estimations. 41 percent of the students who received the invitation signed up for the program (Column 1). Some of these students drop out even before the first meeting, leaving 37 percent of those invited to sign up and taking at least one meeting (Column 2). Of those who were ever invited, 32% take all five meetings (Column 3). The final two columns show that female students are more likely to sign up for the program (conditional on receiving an invitation). This is in line with the findings of Angrist et al. (2009). Conditional on the invitation to sign up, students on average participate in 1.73 meetings. Conditional on initial sign-up, students participate in 4.2 meetings on average. Conditional on participating in at least one meeting, they participate in 4.8 of the 5 meetings on average.

Table 3 shows that female and male mentors differ slightly in how they act as mentors. Female mentors conduct around 0.4 more meetings with their mentees than male mentors (Column 1). This effect is especially pronounced for female mentees, who attend around half a meeting more on average (Column 2). However, the average length of meetings (measured in hours) is a bit lower for female than for male mentors (Column 4). The effect of -0.07 corresponds to meetings held by female mentors being shorter by around 4 minutes, relative to an average meeting length of 39 minutes.

	Meetings					
Dependent Variable:	# attended			Av. length (hours)		
Mentees:	All	Female	Male	All	Female	Male
	(1)	(2)	(3)	(4)	(5)	(6)
Female mentor	0.40**	0.54**	0.28	-0.07***	-0.06**	-0.08**
	(0.17)	(0.26)	(0.23)	(0.02)	(0.03)	(0.03)
Mean dep.	4.66	4.57	4.76	0.65	0.65	0.64
Obs.	128	70	58	128	70	58

Note: This table shows impacts of having a female mentor on mentoring characteristics. The sample includes those who met at least once with their mentors. Columns (1) to (3) use the number of attended meetings as the dependent variable. Columns (4) to (6) use the average meeting length in hours as the dependent variable. All columns control for the number of credits earned in the winter term. Columns (1) and (4) additionally control for mentee gender. Standard errors are robust. * p<0.10, ** p<0.05, *** p<0.01

Table 3: Meetings by Mentor Gender

3.3 Estimation

To evaluate the treatment effects of the peer mentoring program on administrative student outcomes and survey responses, we run linear regressions according to the estimation equation

$$y_i = \alpha + \beta Treatment_i + \gamma_1 Female_i + \gamma_2 CreditsWT_i + \epsilon_i, \quad (1)$$

where y_i is the outcome of interest of student i, $Treatment_i$ is an indicator for (random) assignment to treatment, $Female_i$ is a dummy for the student's gender that takes the value of one if the student is female, and $CreditsWT_i$ is the number of (ECTS) credits earned by the student in the winter term 2019, the first term in which the students were enrolled in the study program. Each of the outcomes is thus regressed on the treatment indicator and the vector of strata variables. We report robust standard errors.

Since not all students in the treatment group took up the offer to receive mentoring services, we additionally run instrumental variable regressions using the randomized treatment assignment as an instrument for actual take-up. The main variable for measuring program take-up is program sign-up (i.e., the first stage can be seen in Column 1 of Table 2). The first stage is expectedly strong, with a Kleibergen-Paap F statistic of around 240.[16]

For several reasons, before implementing the intervention we considered it likely that the treatment would have heterogeneous effects. First, this expectation was based

[16] As can be seen from Table 2, not all students who signed up made use of mentoring services. We therefore also estimated model variants where we used treatment assignment to instrument for actual service use (not shown). Qualitatively, this makes no difference for our results.

on the observation that prior evidence on online education shows more pronounced negative effects among weaker students (e.g., Figlio et al., 2013; Bettinger et al., 2017b). We thus expected treatment effects to differ by the number of credits earned in the first term. In addition, in the baseline, there is a positive correlation between students' high school GPA and the probability to meet the 30 credits target in any term. In Online Appendix B.3, we therefore also show estimates using mentees' high-school GPA as a dimension to study the treatment effect heterogeneity. The results are similar to using credits earned in the first term as the dimension of heterogeneity.

A second observation is that the literature has commonly found male students to suffer more from online relative to in-person teaching (e.g., Figlio et al., 2013; Xu and Jaggars, 2014). At the same time, take-up rates in mentoring programs seem to be higher for female students (e.g., Angrist et al., 2009). Thus, while we expected the effects of mentoring on outcomes among randomly chosen students to be larger for male than for female students, the relative magnitude of effects of having been offered a mentor on outcomes, and the relative effect of mentoring on outcomes conditional on take-up, was ex-ante unclear. We therefore planned to study the effects of our intervention by gender to inform on these questions.

We investigated additional heterogeneities that we already described as less central (and likely not to be reported) in the pre-analysis plan. First, one question is whether the effects of mentoring are larger when being mentored by female than by male mentors. Prior literature has found that interactions between student and instructor gender can matter for teaching effectiveness (e.g., Dee, 2005, 2007; Hoffmann and Oreopoulos, 2009). We study this descriptively in Table A.1 in the Online Appendix, but find no strong results. Second, the pre-analysis plan also specified that we would test if the treatment response of students enrolled at university for the first time differs from students who have been enrolled before. Again, we do not find any strong heterogeneity here (result not reported).

We study the treatment effect heterogeneity by running regressions including an interaction term between the variable capturing the dimension of heterogeneity and the treatment indicator, together with the variable capturing the dimension itself. The strata variables are included as controls. In addition to this (pre-registered) estimation and to provide transparent information from a variety of models, we also split our sample into terciles of initial performance to study the effects by prior performance and into female and male students to study effects by gender. Finally, we also report results from more flexible specifications where we estimate third-order polynomials of interactions between the running variable capturing the heterogeneity and the outcome variable of interest.

4 The Impacts of Peer Mentoring on Administrative Student Outcomes

In this section, we report our results on administrative student outcomes. We start by showing average intent-to-treat and instrumental variable treatment effects on credits for which students registered, credits earned, and GPA in the summer term 2020. We subsequently show the heterogeneity of these effects by prior performance and gender.

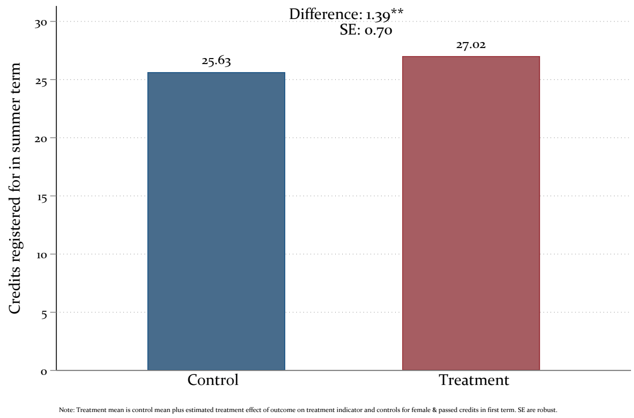
4.1 Impacts on Average Outcomes

Figure 2 shows mean differences between treatment and control group for credits registered for, credits earned, and students' GPA in the summer term.[17] Panel (a) shows the impacts on credits registered for. It shows that students who received a treatment offer register for around 1.4 more credits than students who did not receive the offer of program participation. This difference is around 5% relative to the control group mean and corresponds to around 28% of an additional course. Thus, the treatment offer shifted around every third student to register for an additional course. Students do not pass all courses for which they register. Panel (b) therefore shows differences between treatment and control group in credits earned. Students who received a treatment offer earn around 0.5 (or 10% of a course) more credits than students in the control group. This is an increase of around 3% relative to the control group mean. The difference is statistically insignificant, though. Finally, Panel (c) shows that students' GPA is virtually unaffected, indicating that the change in attempted and earned credits did not come at the expense of worse average grades.

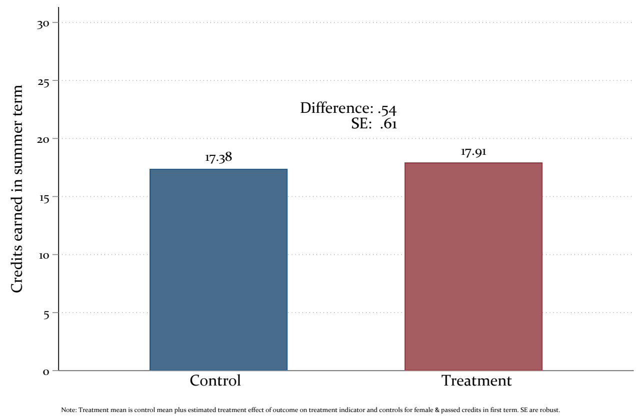
We investigate these results further in Table 4. The odd-numbered columns repeat the ITT estimates displayed in Figure 2. The even-numbered columns show corresponding IV estimates where we use treatment offer as an instrumental variable for program sign-up. All columns control for students' gender and the credits which they earned in their first term of study. Column (2) shows that students who signed up for the treatment register for around 3.4 more credits than those who did not. This corresponds to around 67% of an additional course and 13% of the control group mean. Column (4) shows that they earn around 1.3 credits more, translating to 26% of an additional course and 7% of the control group mean. Thus, students pass around 40% of the additional credits for which they register to take the exam. These results

[17]The figures are analogous to Bergman et al. (2020). In all panels, the control mean is calculated as the students' mean in the control group. Treatment effects, reported in the top center of each panel, are estimated using an OLS regression of the outcome on a treatment indicator, an indicator for students' gender, and students' credits earned in their first term. The treatment mean in the panel is calculated as the control mean plus the estimated treatment effect. Standard errors reported in the panel are robust.

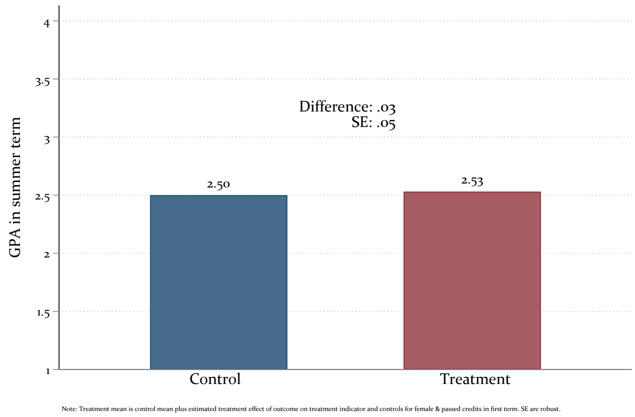
are statistically insignificant, however. Column (6) again shows that students' GPA is largely unaffected.[18]



(a) Credits registered for



(b) Credits earned



(c) GPA

Figure 2: Student Outcomes in the Online Summer Term, by Treatment

Note: This figure shows student outcomes by treatment status. Panel (a) uses the number of credits for which students registered in the summer term 2020 as the outcome measure. Panel (b) uses the number of credits earned in the summer term as outcome measure. Panel (c) uses average GPA (running from 1=worst to 4=best) among earned credits as the outcome measure. In all panels, the control mean is calculated as the students' mean in the control group. Treatment effects, reported in the top center of each panel, are estimated using an OLS regression of the outcome on a treatment indicator, an indicator for students' gender, and students' credits earned in their first term. The treatment mean in the panel is calculated as the control mean plus the estimated treatment effect. Standard errors reported in the panel are robust.

[18]Figure B.1 in the Online Appendix shows outcome distributions by treatment status.

Dependent Variable:	Credits				GPA	
	Registered for		Earned			
	ITT	IV	ITT	IV	ITT	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	1.39**	3.37**	0.54	1.30	0.03	0.07
	(0.70)	(1.69)	(0.61)	(1.47)	(0.05)	(0.11)
Mean dep.	26.33	26.33	17.66	17.66	2.52	2.52
Obs.	691	691	691	691	595	595

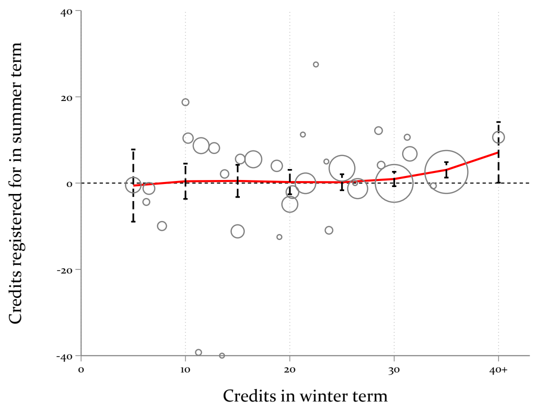
Note: This table shows impacts of peer mentoring on administrative student outcomes using Equation 1. The odd-numbered columns use OLS regressions. The even-numbered columns instrument a dummy for initial program take-up by the (random) treatment assignment variable. Columns (1) and (2) use the number of credits for which students registered in the summer term 2020 as the dependent variable. Columns (3) and (4) use the number of earned credits in the summer term as the dependent variable. Columns (5) and (6) use students' average GPA (running from 1=worst to 4=best) among earned credits in the summer term as the dependent variable. The number of observations differs from Columns (1)-(4) since we have several students who do not earn any credits. Standard errors are robust. * p<0.10, ** p<0.05, *** p<0.01

Table 4: Average Impacts of Online Peer Mentoring on Student Outcomes

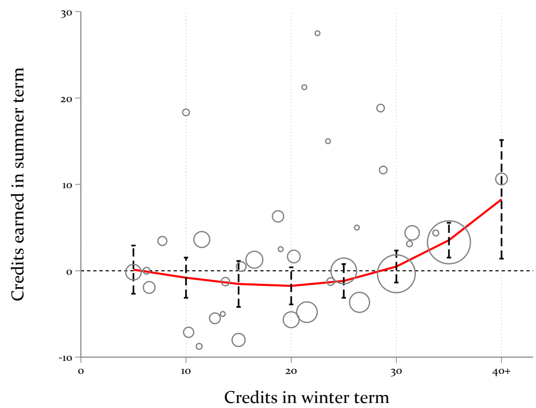
4.2 Impacts by Prior Performance

As outlined before, prior evidence on the effectiveness of online education suggests that its negative effects are more pronounced for weaker students (e.g., Figlio et al., 2013; Bettinger et al., 2017b). We therefore investigate the heterogeneity of these effects in Figure 3. The figure shows the empirical outcome differences (bubbles) and the predicted outcome differences (red solid lines) by credits earned in the winter term, conditional on the strata variables. The underlying model is a third-order polynomial of interactions between the treatment dummy and students' credits earned in the winter term. The spikes indicate 95% confidence intervals.

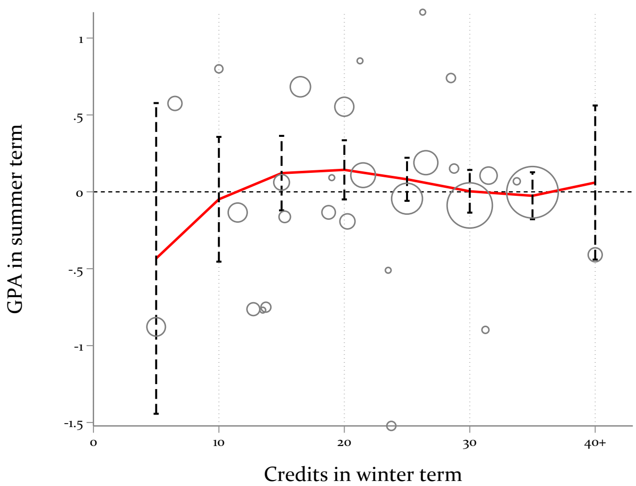
Panel (a) considers as an outcome the credits students registered for. There is hardly any difference between treatment and control group along the lower part of the distribution of credits earned in the winter term, but the difference bends up for students at the upper end of the distribution. The same pattern can be seen in Panel (b), which shows the impacts on credits earned in the summer term. Here, the positive impact of the treatment on students in the upper part of the distribution of past academic achievement is even stronger. Finally, Panel (c) again shows that the effects on GPA are limited, no matter how students fared in the winter term. Overall, Figure 3 demonstrates that the peer mentoring program helped relatively good students in passing more courses. In contrast, the program did not significantly affect students who performed relatively poorly in their first term at university.



(a) Credits registered for



(b) Credits earned



(c) GPA

Figure 3: Treatment Effects by Credits Earned in Winter Term

Note: This figure shows how student outcomes in the summer term relate to students' prior performance as measured by credits earned in the winter term. Panels (a) to (c) display heterogeneous treatment effects (relative to the control group) on credits registered for, credits earned, and students' GPA among earned credits in the summer term 2020, respectively. The bubbles represent empirical differences between treatments, and the red solid lines indicate the treatment effects obtained from the model $y_i = \sum_{j=0}^{3} \beta_j \cdot (x_i)^j + \sum_{j=0}^{3} \gamma_j \cdot (x_i)^j \cdot T_i + u_i$, where y_i is the outcome of interest, x_i is our measure for prior performance, and T_i is an indicator for the treatment group. The spikes indicate 95% confidence intervals (Huber-White standard errors). One student in the sample passed 45 credits in the winter term and is included in group "40+" for better visibility.

We investigate these patterns further in Table 5. In all panels of this table, we report the pre-registered interaction effects between treatment dummy and credits earned along with treatment effects in Column (1). In Columns (2) and (3), we report ITT and IV estimates for students in the bottom tercile of the distribution of credits in the winter term, respectively. In Columns (4) and (5), we report analogous effects for the middle tercile. And in Columns (6) and (7) we report these effects for the top tercile.

Panel (a) shows the impacts on credits for which students registered. The interaction term is insignificant, but again points towards a higher treatment effect for those with more credits in the winter term. This can be seen more clearly in the subsequent columns. While the effects for the bottom tercile are small, they increase in the middle tercile, and in the top tercile we observe significant effects on credits registered for. The instrumental variable estimates in Column (7) suggest that students in the top tercile who received peer mentoring due to our initial offer register for around one more course than similar students in the control group.

Panel (b) repeats this analysis for credits earned. The interaction effect shows that students who passed more credits in the winter term benefit more from the program. The point estimates suggest a positive treatment effect starting at around 23 credits (or five courses) passed in the winter term in which students should have passed six courses. Again, the effects by tercile show that only those in the top tercile benefit from the program. In the highest tercile, treated students pass slightly more than one additional course than similar students in the control group. Panel (c) again shows that we do not see any effects on GPA.

Online Appendix Figure B.2 illustrates this result further. The figure shows the share of students who reach the institutionally recommended goal of having earned 60 credits by the end of the second term, by treatment status and by students' tercile in the distribution of credits earned in the winter term. By construction, the share of students who reach the goal increases across terciles, with none of the students in the control group of the lowest tercile reaching the goal to around 52% of students in the control group in the highest tercile reaching it. While there is no change in the probability of having reached the goal in the lowest tercile, there is a small difference of around 2 percentage points for the middle tercile. For the highest tercile, there is a difference in the probability of reaching the study plan goal after two terms of almost 10 percentage points (p-Value=0.059). This difference amounts to 18% of the control group mean.

Overall, our results therefore paint a consistent picture: Those who fared well in the winter term benefited from the peer mentoring program. In contrast, those who did not earn as many credits in the winter term seem largely unaffected.[19] This is interesting especially because in many investigations of the effectiveness of (peer) mentoring programs in higher education, good students are excluded from the investigation (e.g., Angrist et al., 2009). Note that these results are in line with prior evidence that students perform better when being mentored or taught by persons similar to them (e.g. Dee, 2005; Hoffmann and Oreopoulos, 2009).

4.3 Impacts by Gender

The literature on online education commonly found male students to suffer more from online relative to classroom education than female students (e.g., Figlio et al., 2013; Xu and Jaggars, 2014). We therefore also investigated the treatment effects by gender. We report the results from this analysis in Table 6. The structure of the table is analogous to Table 5. The first column reports the interaction between a dummy for female student gender. Columns (2) and (3) report effects for female students. Columns (4) and (5) report results for male students.

Panel (a) again shows results for credits registered for. The first column shows a positive treatment effect for men, who register for around 2.7 more credits (more than half an additional course) when offered treatment. The interaction effect is negative and of around the same magnitude, suggesting that female students do not benefit from the program. This can also be seen in the remaining columns. While we do not see any impact for female students, male students register for more than one additional course in the instrumental variable estimates (Column 5).

Panel (b) shows a similar pattern for credits earned. The results are again attenuated, however. Column (1) shows an effect of around 0.9 more credits earned by male students, with zero effects for female students. Columns (2) and (3) again show no effects for female students. In Columns (4) and (5), we see a positive but insignificant effect on male students. Male students who take up the offer pass around 2.5 credits (half a course) more than students who do not. This is an increase of around 16% relative to the control group mean and suggests that male students pass around a third of the additional credits that they register for. Panel (c) again shows no effects on GPA.[20]

[19]For a similar analysis using high-school GPA as measure of prior performance, see Online Appendix B.3. The results are slightly more U-shaped, suggesting that weak students may also benefit somewhat from the program.

[20]In Online Appendix B.4, Figure B.4 shows bar charts for these results. Figures B.5 and B.6 show heterogeneous treatment effects by credits earned in the winter term, by gender. Again, both female and male students benefit more when they passed more credits in the winter term. However, this pattern is much more pronounced for male students.

Panel A: Credits registered for							
		Lowest tercile		Middle tercile		Highest tercile	
	ITT	ITT	IV	ITT	IV	ITT	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	-0.55	0.25	0.62	1.86	3.52	1.91***	5.32***
	(2.92)	(1.77)	(4.30)	(1.27)	(2.41)	(0.69)	(1.97)
Treatment · credits (WT)	0.08						
	(0.10)						
Mean dep.	26.33	23.31	23.31	27.47	27.47	27.73	27.73
Obs.	691	209	209	156	156	326	326
Panel B: Credits earned							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	-4.13***	-0.89	-2.21	-0.91	-1.73	2.14**	5.95**
	(1.57)	(1.06)	(2.67)	(1.51)	(2.83)	(0.83)	(2.33)
Treatment · credits (WT)	0.18***						
	(0.06)						
Mean dep.	17.66	7.93	7.93	16.94	16.94	24.24	24.24
Obs.	691	209	209	156	156	326	326
Panel C: GPA							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.14	0.17	0.31	0.05	0.10	-0.02	-0.06
	(0.21)	(0.12)	(0.22)	(0.09)	(0.16)	(0.07)	(0.18)
Treatment · credits (WT)	-0.00						
	(0.01)						
Mean dep.	2.52	2.01	2.01	2.25	2.25	2.84	2.84
Obs.	595	129	129	144	144	322	322

Note: This table shows impacts of peer mentoring on administrative student outcomes by prior performance adapting equation 1. In each panel, the first column uses the baseline sample. After the first column, the even-numbered columns use OLS regressions. The odd-numbered columns instrument a dummy for initial program take-up by the (random) treatment assignment variable. Columns (2) and (3) use students in lowest tercile of the distribution of credits earned in the winter term, the first term in which the students studied. Columns (4) and (5) use those in the middle tercile and Columns (6) and (7) those in the highest tercile. The terciles are differentially large because earned credits are not continuous. The regressions in Panel (a) use the number of credits for which students registered in the summer term 2020 as the dependent variable. The regressions in Panel (b) use the number of earned credits in the summer term as the dependent variable. The regressions in Panel (c) use students' average GPA (running from 1=worst to 4=best) among earned credits in the summer term as the dependent variable. The number of observations differs from Panels (a) and (b) since we have several students who do not earn any credits. Standard errors are robust. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table 5: Treatment Effects by Credits Earned in Winter Term

Panel A: Credits registered for					
		Female		Male	
	ITT	ITT	IV	ITT	IV
	(1)	(2)	(3)	(4)	(5)
Treatment	2.67***	-0.09	-0.19	2.65***	7.28***
	(0.99)	(0.98)	(2.10)	(0.99)	(2.72)
Treatment · female	-2.73*				
	(1.40)				
Mean dep.	26.33	27.63	27.63	25.19	25.19
Obs.	691	324	324	367	367
Panel B: Credits earned					
	(1)	(2)	(3)	(4)	(5)
Treatment	0.91	0.11	0.23	0.90	2.48
	(0.83)	(0.90)	(1.92)	(0.83)	(2.26)
Treatment · female	-0.79				
	(1.22)				
Mean dep.	17.66	19.14	19.14	16.36	16.36
Obs.	691	324	324	367	367
Panel C: GPA					
	(1)	(2)	(3)	(4)	(5)
Treatment	0.03	0.04	0.09	0.03	0.07
	(0.07)	(0.08)	(0.15)	(0.07)	(0.17)
Treatment · female	0.01				
	(0.10)				
Mean dep.	2.52	2.54	2.54	2.49	2.49
Obs.	595	291	291	304	304

Note: This table shows impacts of peer mentoring on administrative student outcomes by gender, adapting equation 1. In each panel, the first column uses the baseline sample. After the first column, the even-numbered columns use OLS regressions. The odd-numbered columns instrument a dummy for initial program take-up by the (random) treatment assignment variable. Columns (2) and (3) use female students only. Columns (4) and (5) use male students only. The regressions in Panel (a) use the number of credits for which students registered in the summer term 2020 as the dependent variable. The regressions in Panel (b) use the number of earned credits in the summer term as the dependent variable. The regressions in Panel (c) use students' average GPA (running from 1=worst to 4=best) among earned credits in the summer term as the dependent variable. The number of observations differs from Panels (a) and (b) since we have several students who do not earn any credits. Standard errors are robust. * p<0.10, ** p<0.05, *** p<0.01

Table 6: Treatment Effects by Mentee Gender

These results stand in contrast to prior evidence on the effectiveness of mentoring programs. Most importantly, Angrist et al. (2009) find that an in-person program that combined academic counseling with financial incentives did positively affect female college students, while having no effects at all on male students. While our results on the gender heterogeneity of the treatment effect on credits earned are imprecisely estimated, the patterns in our data do not suggest a similar heterogeneity as in Angrist et al. (2009). If anything, in our context male students seem to benefit more from the peer mentoring program compared to female students. This may be explained by the online teaching environment which has been shown to particularly impair the performance of male students (e.g., Figlio et al., 2013).

4.4 Additional Analyses in the Online Appendix

We add to the pre-registered heterogeneity analyses in the Online Appendix. In Online Appendix B.5, we follow Abadie et al. (2018) and Ferwerda (2014) and estimate treatment effects using endogenous stratification approaches. In line with the analysis by prior performance, students in the upper tercile of the distribution of predicted outcomes in the summer term seem to benefit most from the program. In Online Appendix B.6, we show results by students' region of origin. We do not find strong heterogeneities here. If anything, students who come from the region where the university is located seem to benefit more from the program.

In Online Appendix B.7, we also provide results from a sort of "value-added" analysis where we regress students' outcomes in the summer term on their performance in the winter term, their observable characteristics, and a mentor dummy. We label the estimated mentor fixed effect in each outcome dimension the mentor's "value-added". We find that mentors differ substantially in their value-added. The value-added estimates on earned credits range from minus 3 to plus 7 credits conditional on students' observables. Thus, mentor differences seem to matter. We also test whether the value-added estimates are correlated across different outcomes measures. While mentors' value-added on credits earned and credits registered for are strongly correlated (ρ=0.56, $p < 0.05$), both measures are not substantially correlated with value-added on GPA. We caution that each mentor only advises up to 10 mentees, thus leading to substantial noise in the value-added estimates. However, the mentors were randomly assigned to mentees, such that we do not have the problem of endogenous sorting of students to mentors common in the literature on teacher value-added. With only 15 mentors, we cannot credibly identify the sources of mentors' performance differences.

5 Survey Evidence on Students' Study Behavior and Views

To better understand the channels through which the treatment effects operate, we now turn to our survey on study behavior and students' views on the department and on online teaching in the summer term and generally. We start by running the same balancing checks that we conducted on the overall sample on the sample of survey respondents. We also study the selectivity in survey participation by means of mean-comparison tests between survey participants and non-participants. Table 7 shows our results. As can be seen from the table, students who did participate in the survey differ slightly from students who did not participate (Columns 1-4). Participants are somewhat younger, more likely to be female, have better high-school GPA, have earned more credits in the winter term, and are more likely to be part-time students. Importantly, the likelihood of completing the survey seems unrelated to treatment assignment. Columns (5) to (8) show that within the sample of survey participants, the treatment and control groups were balanced across all individual characteristics.

Figure 4 shows results from ITT and IV estimations using equation 1 and instrumenting take-up by treatment assignment, respectively. For better visibility, we show the treatment effects only and provide 90% confidence intervals. All corresponding tables can be found in Online Appendix C. All dependent variables are survey responses to questions, measured on a five-point Likert scale where higher values indicate higher agreement with the question. The specific questions can be found in Online Appendix C.1.

Panel (a) shows treatment effects on students' assessment of their own study motivation and study behavior in the summer term 2020. These are outcomes that the mentoring program specifically targeted. The first two rows show positive impacts on students' self-reported motivation. The estimated treatment effect in the instrumental variables estimation amounts to around half a point on a five-point Likert scale or around 18% relative to the mean of the dependent variable. The next two rows show significant effects on students' response to the question whether they managed to study continuously throughout the summer term. The subsequent two rows show smaller effects on students' response to the question whether they think they prepared for exams in time. The final two rows again show significant effects on students' response to the question whether they think they provided enough effort to reach their goals. To complement these results, we also estimate average standardized effects analogous to Kling et al. (2004) and Clingingsmith et al. (2009) in Online Appendix Table C.5. This part of the survey shows an average standardized treatment effect of around 0.16 standard deviations (p-value = 0.048). In Online Appendix C.4, we show the heterogeneity of these effects by credits earned in the winter term, analogously to Section 4.2. While the results are more noisy, the overall pattern of this heterogeneity is similar to the one in Figure 3.

Panel (b) shows that the treatment did not shift views on departmental services. The aspects include students services, communication by the department, whether there is a clear contact person from the department, as well as students' views on whether the department cares for their success or takes their concerns seriously. The most pronounced effect is for students' feeling whether the department cares for their success, with point estimates of around 0.2 points on a five-point Likert scale or 7% relative to the mean of the dependent variable. This is not significantly different from zero, however. The associated average standardized effect is 0.03 standard deviations (p-value = 0.65).

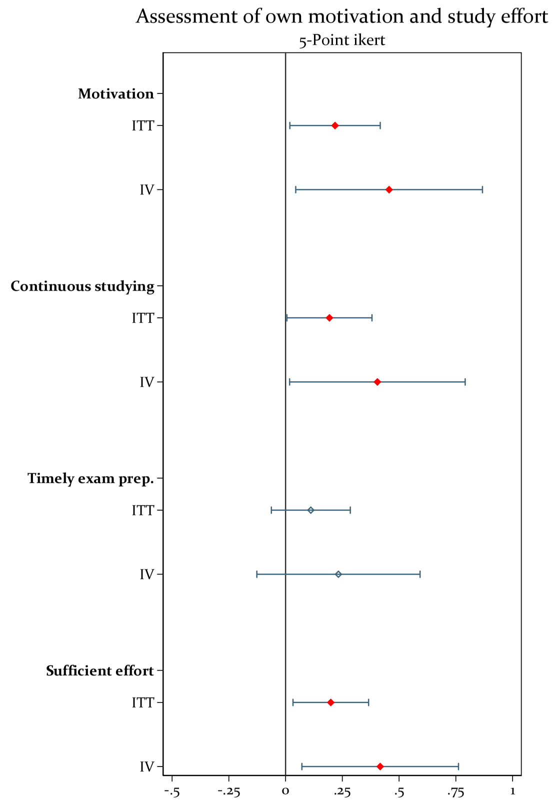
Panel (c) then shows results on students' views of online teaching in the summer term 2020 and more generally. The first four rows show results on students' satisfaction with the departments' online teaching content and technical implementation in the summer term. Students' views on these aspects of the summer term seem unaffected by the treatment. The next two rows show students' response to the question whether they frequently interacted with other students in some form. The null result is interesting since it shows that the program did not merely substitute for interactions among students. The final two rows show students' views on online teaching more generally. We asked students whether they feel that online teaching can work in principle and whether online teaching should play a large role in the future. Both sets of results are insignificantly different from zero. However, the response to students' views on whether online teaching should play an important role in the future shows a point estimate in the IV regressions of 0.2 points on a five-point Likert scale or 6% relative to the mean of the dependent variable. The associated standardized treatment effect of this part of the survey is 0.02 (p-value = 0.72).

In the survey, we additionally elicited students' expectations of the likelihood of completing their studies in time and the number of planned credits. The results are noisy and show no difference between treatment and control group (not shown). The contrast to the results using administrative student outcomes shown in Section 4 highlight the importance of analyzing actual administrative data instead of students' survey responses.

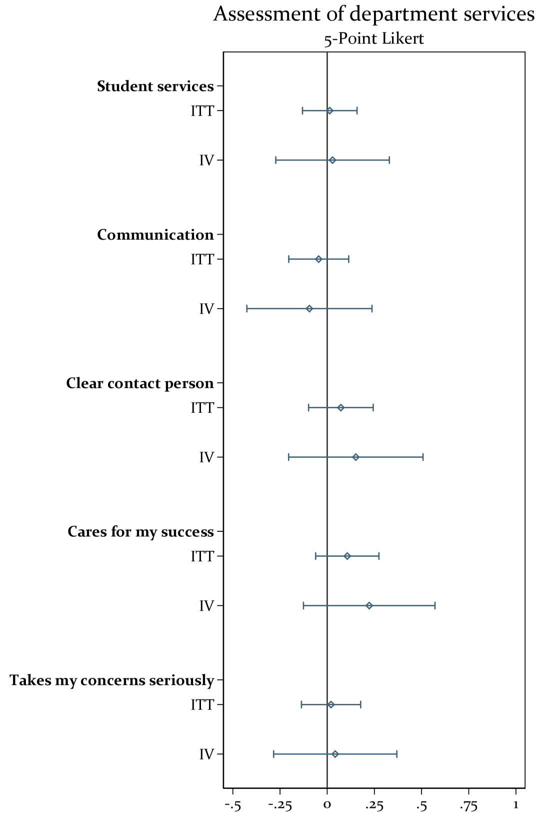
	Survey participation				Within survey			
	Non-participants	Participants	Difference	Std.diff.	Control	Treatment	Difference	Std.diff.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment group	0.51	0.49	-0.01	-0.02				
	(0.50)	(0.50)	(0.04)					
Female	0.41	0.51	0.10***	0.15	0.52	0.50	-0.02	-0.03
	(0.49)	(0.50)	(0.04)		(0.50)	(0.50)	(0.05)	
Age	21.57	21.06	-0.50**	-0.13	20.93	21.19	0.26	0.09
	(3.07)	(2.15)	(0.20)		(1.79)	(2.46)	(0.21)	
High-school GPA	2.24	2.47	0.23***	0.28	2.47	2.47	-0.01	-0.01
	(0.55)	(0.60)	(0.05)		(0.55)	(0.64)	(0.06)	
Top-tier high-school type	0.77	0.74	-0.03	-0.05	0.77	0.70	-0.07	-0.11
	(0.42)	(0.44)	(0.03)		(0.42)	(0.46)	(0.04)	
Foreign univ. entrance exam	0.08	0.07	-0.02	-0.04	0.07	0.06	-0.01	-0.04
	(0.28)	(0.25)	(0.02)		(0.26)	(0.24)	(0.02)	
Earned credits in first term	21.51	27.90	6.39***	0.52	28.00	27.80	-0.20	-0.02
	(9.86)	(7.46)	(0.66)		(7.55)	(7.39)	(0.74)	
First enrollment	0.64	0.66	0.02	0.02	0.62	0.70	0.07	0.11
	(0.48)	(0.47)	(0.04)		(0.49)	(0.46)	(0.05)	
Part-time student	0.05	0.11	0.07***	0.18	0.12	0.11	-0.02	-0.04
	(0.21)	(0.32)	(0.02)		(0.33)	(0.31)	(0.03)	
Obs.	287	404	691	691	205	199	404	404

Table 7: Sorting into Survey Participation

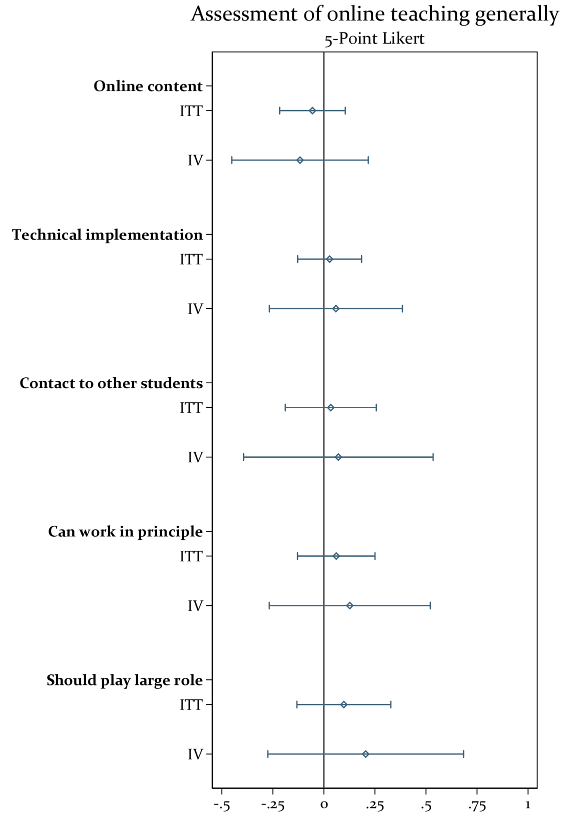
Note: This table shows selection into survey participation. The first four columns show means administrative student data of participants and non-participants along with differences between both groups. The next four columns show means and differences in administrative student data by initial treatment assignment among survey participants. We estimated whether the differences between groups are statistically significant using t-tests in Columns (3) and (7) and using standardized differences in Columns (4) and (8). * $p<0.10$, ** $p<0.05$, *** $p<0.01$



(a) Assessment of study behavior



(b) Assessment of department services



(c) Assessment of online teaching

Figure 4: Impacts of Mentoring on Survey Outcomes

Note: This figure shows impacts of peer mentoring on survey responses adapting Equation 1. For each question, the first row uses OLS regressions (labeled “ITT”). The second row uses (random) treatment assignment as an instrument for initial program take-up (labeled “IV”). For the full set of survey questions, please see Online Appendix C.1. Diamonds indicate the point estimates, bars the associated confidence bounds. Full (red) diamonds indicate significance at the ten percent level. Hollow diamonds indicate non-significance. The corresponding tables can be found in Online Appendix C.2. Standard errors are robust.

Overall, our results show that the treatment improved students' motivation and study behavior, which is exactly what the peer mentoring program intended to shift. We thus view these survey outcomes as evidence on the mechanism through which the mentoring program worked. The program seems to have affected students' behavior and well-being during the online summer term which led to more exam registrations and somewhat better student outcomes. In contrast, views on departmental services or online teaching were not shifted significantly by the treatment. If anything, we see that students feel somewhat more positively towards a future role of online teaching and that the department cares for their success.

6 Conclusion

This paper presents the first evidence on the potential role of remote peer mentoring programs in online higher education. We conducted a field experiment that provided first year students with an online mentor from a more advanced term. The structured one-on-one mentoring program focused on study behavior, (online education) study skills, and students' self-organization, some of the most common issues in online teaching. For our experiment, we leveraged the COVID-19-induced switch to online teaching at a large German public university, where the entire summer term was conducted online.

Our peer mentoring program increased exam registrations and, somewhat attenuated, earned credits among affected students. In contrast to prior research, our results are more pronounced for previously good students. If anything, the program was also more effective for men. The program strongly affected students' behavior and well-being during the online term. Our results thus provide the first evidence on the effectiveness of a low-cost intervention such as peer mentoring to improve student outcomes and student well-being in online higher education.

Given the cumulative nature of human capital accumulation, our results on students' well-being and behavior may suggest that a more permanent peer mentoring program may improve student outcomes even more. However, it may also be the case that the novel nature of the mentoring program, combined with the unusual situation of a rapid switch to online teaching, exacerbate any treatment effects. The potential role of peer mentoring programs in online education thus remains a valuable topic for future research.

References

Abadie, A., M. M. Chingos, and M. R. West (2018): "Endogenous Stratification in Randomized Experiments," *Review of Economics and Statistics*, 100, 567–580.

Alpert, W. T., K. A. Couch, and O. R. Harmon (2016): "A Randomized Assessment of Online Learning," *American Economic Review: Papers & Proceedings*, 106, 378–382.

Angrist, J., D. Lang, and P. Oreopoulos (2009): "Incentives and Services for College Achievement: Evidence from a Randomized Trial," *American Economic Journal: Applied Economics*, 1, 136–163.

Aucejo, E. M., J. French, M. P. U. Araya, and B. Zafar (2020): "The Impact of COVID-19 on Student Experiences and Expectations: Evidence from a Survey," *Journal of Public Economics*, 191, 104271.

Bacher-Hicks, A., J. Goodman, and C. Mulhern (2020): "Inequality in Household Adaptation to Schooling Shocks: Covid-Induced Online Learning Engagement in Real Time," *Journal of Public Economics*.

Banerjee, A. V. and E. Duflo (2014): "(Dis)Organization and Success in an Economics MOOC," *American Economic Review: Papers & Proceedings*, 104, 514–518.

Bergman, P., R. Chetty, S. DeLuca, N. Hendren, L. F. Katz, and C. Palmer (2020): "Creating Moves to Opportunity: Experimental Evidence on Barriers to Neighborhood Choice," *NBER Working Paper No. 26164*.

Bettinger, E. P. and R. B. Baker (2014): "The Effects of Student Coaching: An Evaluation of a Randomized Experiment in Student Advising," *Educational Evaluation and Policy Analysis*, 36, 3–19.

Bettinger, E. P., C. Doss, S. Loeb, A. Rogers, and E. S. Taylor (2017a): "The Effects of Class Size in Online College Courses: Experimental Evidence," *Economics of Education Review*, 58, 68–85.

Bettinger, E. P., L. Fox, S. Loeb, and E. S. Taylor (2017b): "Virtual Classrooms: How Online College Courses Affect Student Success," *American Economic Review*, 107, 2855–2875.

Bettinger, E. P., B. T. Long, P. Oreopoulos, and L. Sanbonmatsu (2012): "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block Fafsa Experiment," *Quarterly Journal of Economics*, 127, 1205–1242.

Bird, K. A., B. L. Castleman, and G. Lohner (2020): "Negative Impacts From the Shift to Online Learning During the COVID-19 Crisis: Evidence from a Statewide Community College System," EdWorkingPaper 20-299.

Brown, B. W. and C. F. Liedholm (2002): "Can Web Courses Replace the Classroom in Principles of Microeconomics?" *American Economic Review: Papers and Proceedings*, 92, 444–448.

Carlana, M. and E. L. Ferrara (2020): "Apart but Connected: Online Tutoring to Mitigate the Impact of COVID-19 on Educational Inequality," Mimeo.

Castleman, B. L. and L. C. Page (2015): "Summer Nudging: Can Personalized Text Messages and Peer Mentor Outreach Increase College Going Among Low-Income High School Graduates?" *Journal of Economic Behavior & Organization*, 115, 144–160.

Clingingsmith, D., A. I. Khwaja, and M. Kremer (2009): "Estimating the Impact of the Hajj: Religion and Tolerance in Islam's Global Gathering," *Quarterly Journal of Economics*, 124, 1133–1170.

Dee, T. S. (2005): "A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?" *American Economic Review: Papers and Proceedings*, 95, 158–165.

——— (2007): "Teachers and the Gender Gaps in Student Achievement," *Journal of Human Resources*, 42, 528–554.

Deming, D. J., C. Goldin, L. F. Katz, and N. Yuchtman (2015): "Can Online Learning Bend the Higher Education Cost Curve?" *American Economic Review: Papers and Proceedings*, 105, 496–501.

Deming, D. J., M. Lovenheim, and R. W. Patterson (2018): "The Competitive Effects of Online Education," in *Productivity in Higher Education*, ed. by C. Hoxby and K. Stange.

Deming, D. J., N. Yuchtman, A. Abulafi, C. Goldin, and L. F. Katz (2016): "The Value of Postsecondary Credentials in the Labor Market: An Experimental Study," *American Economic Review*, 106, 778–806.

Ferwerda, J. (2014): "ESTRAT: Stata module to perform Endogenous Stratification for Randomized Experiments," *Statistical Software Components S457801*.

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

Goodman, J., J. Melkers, and A. Pallais (2019): "Can Online Delivery Increase Access to Education?" *Journal of Labor Economics*, 37, 1–34.

Grewenig, E., P. Lergetporer, K. Werner, L. Woessmann, and L. Zierow (2020): "COVID-19 and Educational Inequality: How School Closures Affect Low- and High-Achieving Students," *CESifo Working Paper No. 8648*.

Himmler, O., R. Jaeckle, and P. Weinschenk (2019): "Soft Commitments, Reminders, and Academic Performance," *American Economic Journal: Applied Economics*, 11, 114–142.

Hoffmann, F. and P. Oreopoulos (2009): "A Professor Like Me: The Influence of Instructor Gender on College Achievement," *Journal of Human Resources*, 44, 479–494.

Joyce, T., S. Crockett, D. A. Jaeger, O. Altindag, and S. D. O'Connell (2015): "Does Classroom Time Matter?" *Economics of Education Review*, 46, 64–77.

Kling, J. R., J. B. Liebman, L. F. Katz, and L. Sanbonmatsu (2004): "Moving to Opportunity and Tranquility: Neighborhood Effects on Adult Economic Self-Sufficiency and Health from a Randomized Housing Voucher Experiment," *KSG Working Paper No. RWP04-035*.

Lavy, V. and A. Schlosser (2005): "Targeted Remedial Education for Underperforming Teenagers: Costs and Benefits," *Journal of Labor Economics*, 23, 839–874.

Lyle, D. S. and J. Z. Smith (2014): "The Effect of High-Performing Mentors on Junior Officer Promotion in the US Army," *Journal of Labor Economics*, 32, 229–258.

McPherson, M. S. and L. S. Bacow (2015): "Online Higher Education: Beyond the Hype Cycle," *Journal of Economic Perspectives*, 29, 135–154.

Oreopoulos, P., R. S. Brown, and A. M. Lavecchia (2017): "Pathways to Education: An Integrated Approach to Helping At-Risk High School Students," *Journal of Political Economy*, 125, 947–984.

Rodriguez-Planas, N. (2012): "Longer-Term Impacts of Mentoring, Educational Services, and Learning Incentives: Evidence from a Randomized Trial in the United States," *American Economic Journal: Applied Economics*, 4, 121–139.

Xu, D. and S. S. Jaggars (2014): "Performance Gaps between Online and Face-to-Face Courses: Differences across Types of Students and Academic Subject Areas," *The Journal of Higher Education*, 85, 633–659.

APPENDIX: FOR ONLINE PUBLICATION ONLY UNLESS REQUESTED OTHERWISE

A Additional Information on the Mentoring Program

A.1 Mentoring Program Structure: Examples

In this subsection, we show some examples of the content of the mentoring program. In Figure A.1, we show a screenshot of the weekly study plan that we provide as an example plan for mentors. In Figure A.2, we show a screenshot of an actual weekly study plan handed in by a mentee. In Figure A.3 we show a screenshot of a brief learning model with learning techniques that we instruct mentors to discuss with their mentees.

KURSPLAN					ZEITPLANBEGINN 9:00	ZEITINTERVALL 30 MIN	Kursliste
UHRZEIT	SONNTAG	MONTAG	DIENSTAG	MITTWOCH	DONNERSTAG	FREITAG	SAMSTAG
9:00		MTH-113		MTH-113		MTH-113	
9:30							
10:00							
10:30							
11:00			HPE 295		HPE 295		
11:30							
12:00							
12:30							
13:00		WR 121		WR 121			
13:30	0		0		0		0
14:00							

Figure A.1: Example: Input for weekly Study Plan

Fächer		KW 17 20.04 - 26.04 Woche 1	KW 18 27.04 - 03.05 Woche 2	KW 19 04.05 -10.05 Woche 3	KW 20 11.05-17.05 Woche 4	KW 21 18.05-24.05 Woche 5
Finanzmathe	Vorlesung	Vorlesung 1	Vorlesung 2	Vorlesung 3	Vorlesung 4	Vorlesung 5
	Tutorium		Tutorium 1	Tutorium 2	Tutorium 3	Tutorium 4
Makro	Vorlesung	Vorlesung 1	Vorlesung 2	Vorlesung 3	Vorlesung 4	Vorlesung 5
	Übung		Übung 1	Übung 2		
	Online Selbsttest		Online Selbsttest 1	Online Selbsttest		
	Klausuren Kurs (1x)					
	Klausuren Tutorien (2x)					
Mikro	Vorlesung	Vorlesung 1	Vorlesung 2	Vorlesung 3	Vorlesung 4	Vorlesung 5
	Übung		Übung 1	Übung 2	Übung 3	Übung 4
	Tutorium		Tutorium 1	Tutorium 2	Tutorium 3	Tutorium 4
	Test				Test 1	
	Klausurenkurs (1x)					
Absatz	Vorlesung	Vorlesung 1	Vorlesung 2	Vorlesung 3	Vorlesung 4	Vorlesung 5
	Übung					Übung 1
	Tutorium					
	(Virtuelle Sprechstunde) freiwiliig					**20.05** 13.15
Jahresabschluss	Vorlesung (Modul)	Modul 1	Modul 2	Modul 3		
	Übung					
	Assignment (Bonus)			AS: 1 - 10.05.20		AS: 2 - 24.05.20

Figure A.2: Example: Actual Weekly Study Plan

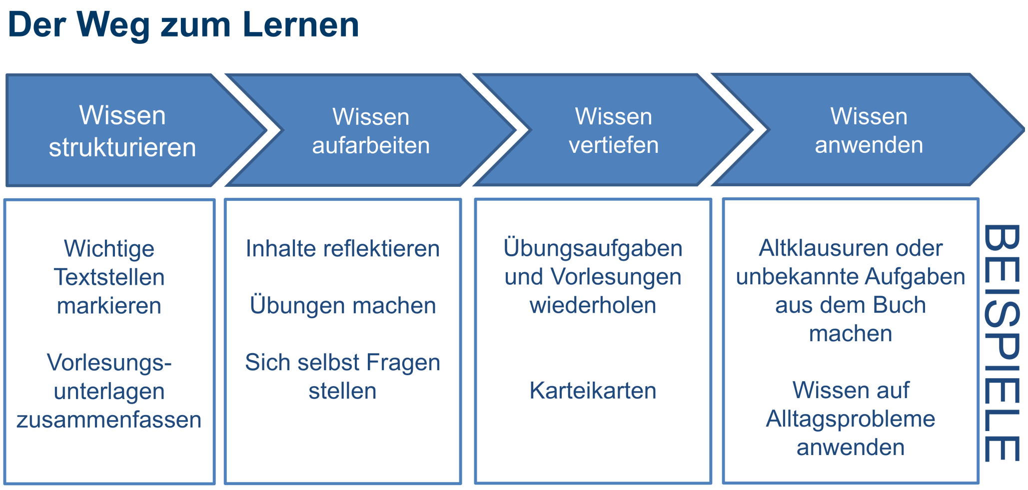


Figure A.3: Example: Input on How to Study Effectively

A.2 Mentoring Differences by Mentor Gender

In this subsection, we provide descriptive evidence on the efficiency of mentoring by mentor gender. We use only those students who signed up for the program, an endogenous outcome itself. Table A.1 shows impacts of having a female mentor on credits for which students registered in the summer term, credits earned, and GPA. If anything, female mentors seem to be a bit more efficient. This is more pronounced for male mentees. The difference between all mentees' outcomes by mentor gender are insignificant, however.

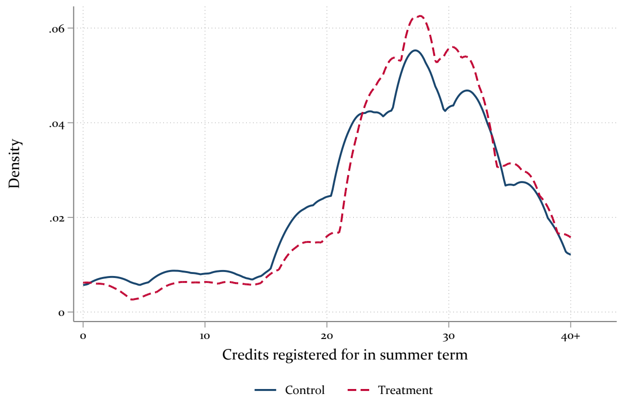
Table A.1: Descriptive Results on Mentoring Effectiveness by Mentor Gender

Dep. Var.:	Credits						GPA		
	Registered for			Earned					
Mentees:	All	Female	Male	All	Female	Male	All	Female	Male
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female mentor	0.75	-0.02	1.98	1.42	1.04	2.34	0.09	0.20	-0.06
	(1.35)	(1.93)	(1.64)	(1.46)	(1.92)	(2.19)	(0.12)	(0.17)	(0.15)
Mean dep.	29.21	29.50	28.86	19.84	20.73	18.78	2.61	2.60	2.62
Obs.	128	70	58	128	70	58	120	65	55

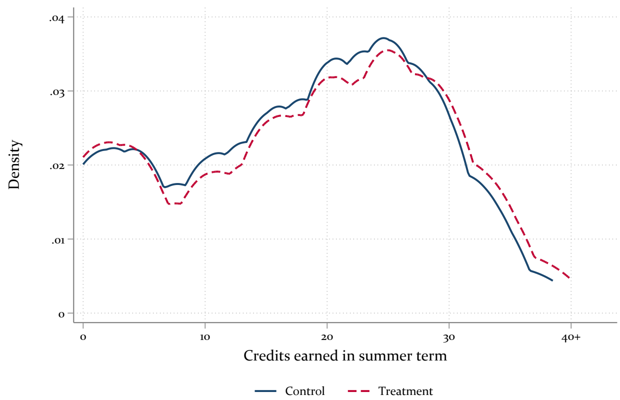
Note: This table shows impacts of having a female mentor on outcomes. The sample includes those who met at least once with their mentors. Columns (1) to (3) use the number of credits for which students registered in the summer term 2020 as the dependent variable. Columns (4) to (6) use the number of credits earned as the dependent variable. Columns (7) to (9) use students' GPA among earned credits as the dependent variable (running from 1=worst to 4=best). All columns control for the number of credits earned in the winter term and for mentee gender. The number of observations differs for these results since not all students have credits earned. Standard errors are robust. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

B Additional Results for Administrative Student Outcomes

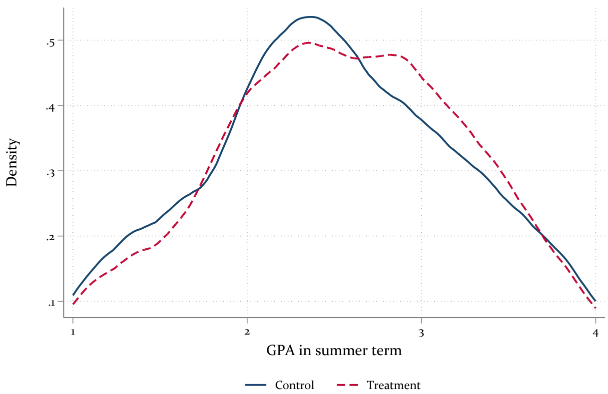
B.1 Distribution of Outcomes by Treatment Status



(a) Credits registered for



(b) Credits earned



(c) GPA

Figure B.1: Kernel Density Plots by Treatment Status

Note: This figure shows unadjusted Kernel density plots by treatment status. Panel (a) uses the number of registered credits in the summer term 2020 as measure. Panel (b) uses earned credits in the same term as measure. Panel (c) uses GPA among earned credits (running from 1=worst to 4=best) as measure. For better visibility, we assign students with above 40 credits to the group "40+" for both credits registered for and credits earned.

B.2 Effects on Reaching First Year Goal

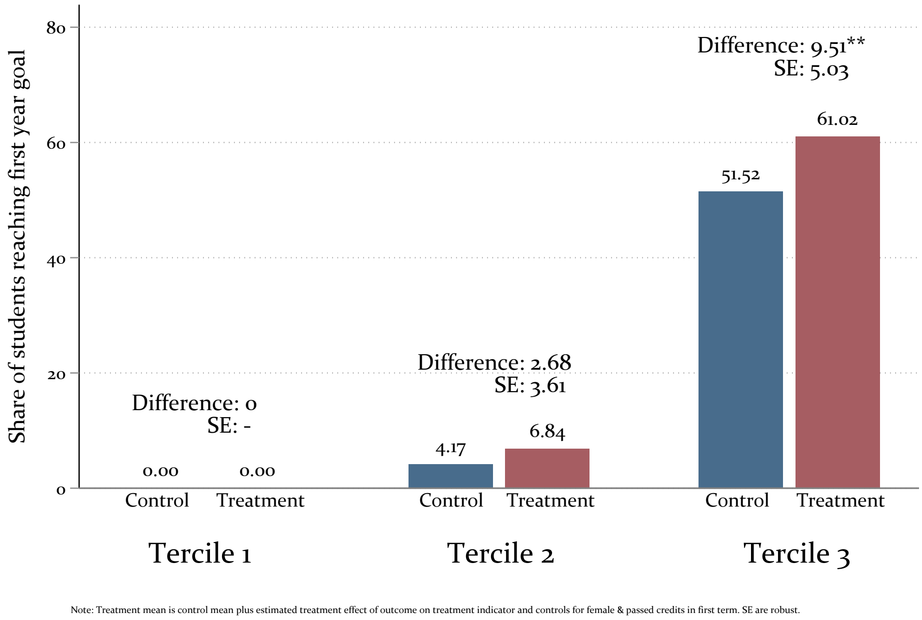
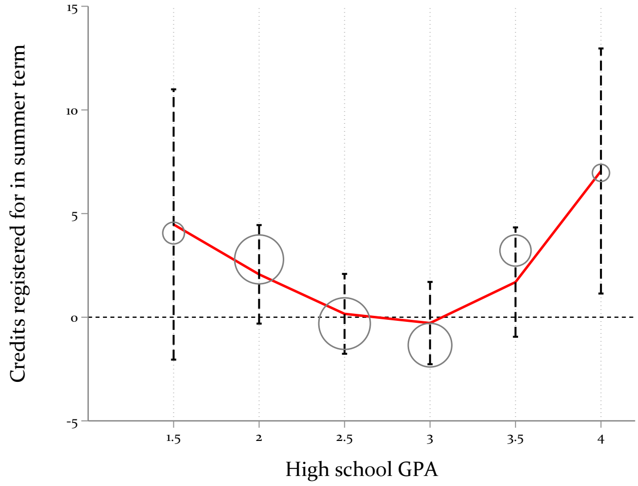


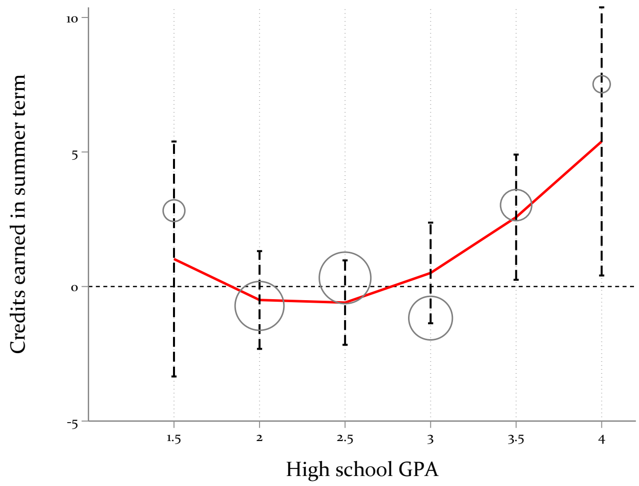
Figure B.2: Treatment Effects by Tercile of Credits Earned in Winter Term

Note: This figure shows the share of students that reached the first year goal of accumulating 60 credits by students' prior performance as measured by their tercile in the distribution of credits earned in the winter term. The control mean is calculated as the students' mean in the control group. Treatment effects, reported in the top center of each comparison, are estimated using an OLS regression of the outcome on a treatment indicator, an indicator for students' gender, and students' credits earned in their first term. The treatment mean is calculated as the control mean plus the estimated treatment effect. Standard errors reported are robust. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

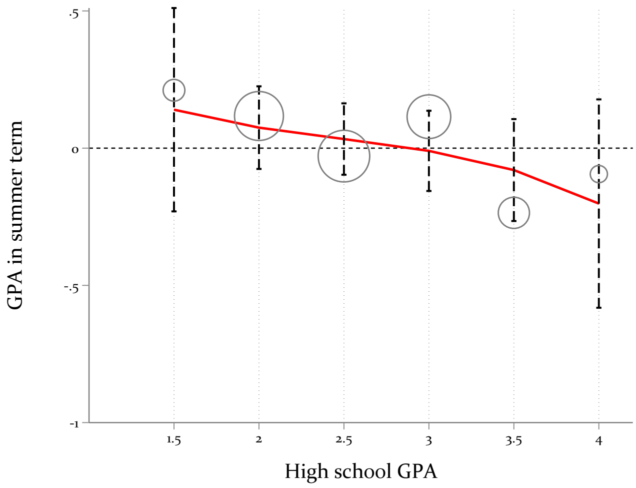
B.3 Heterogeneity by High School GPA



(a) Credits registered for



(b) Credits earned



(c) GPA

Figure B.3: Treatment Effects by Binned High School GPA

Note: This figure shows how student outcomes in the summer term relate to students' prior performance as measured by students' binned high-school GPA. We bin high-school GPA (running from 1=worst to 4=best) into bins of 0.5. Panels (a) to (c) display heterogeneous treatment effects (relative to the control group) on credits registered for, credits earned, and students' GPA among earned credits in the summer term 2020, respectively. The bubbles represent empirical differences between treatments, and the red solid lines indicate the treatment effects obtained from the model $y_i = \sum_{j=0}^{3} \beta_j \cdot (x_i)^j + \sum_{j=0}^{3} \gamma_j \cdot (x_i)^j \cdot T_i + u_i$, where y_i is the outcome of interest, x_i is our measure of prior performance, and T_i is an indicator for the treatment group. The spikes indicate 95% confidence bands (Huber-White standard errors). We also drop 11 observations where we do not have information on students' high-school GPA.

Table B.1: Treatment Effects by High School GPA

Panel A: Credits registered for

		Lowest tercile		Middle tercile		Highest tercile	
	ITT	ITT	IV	ITT	IV	ITT	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	2.82	2.99**	6.44**	-0.77	-2.11	1.41	3.51
	(2.33)	(1.41)	(3.01)	(1.28)	(3.53)	(0.99)	(2.43)
Treatment · H.S. GPA	-0.64						
	(0.88)						
Mean dep.	26.37	25.67	25.67	25.27	25.27	27.93	27.93
Obs.	680	224	224	208	208	248	248

Panel B: Credits earned

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	-9.94***	-0.01	-0.01	-0.81	-2.21	2.09**	5.21**
	(1.71)	(1.01)	(2.17)	(1.11)	(3.06)	(0.98)	(2.41)
Treatment · H.S. GPA	4.37***						
	(0.67)						
Mean dep.	17.68	11.90	11.90	16.62	16.62	23.80	23.80
Obs.	680	224	224	208	208	248	248

Panel C: GPA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	-0.67***	0.12	0.23	-0.07	-0.16	0.06	0.13
	(0.16)	(0.08)	(0.17)	(0.09)	(0.22)	(0.08)	(0.18)
Treatment · H.S. GPA	0.29***						
	(0.06)						
Mean dep.	2.52	2.09	2.09	2.48	2.48	2.87	2.87
Obs.	586	175	175	180	180	231	231

Note: This table shows impacts of peer mentoring on administrative student outcomes by prior performance adapting equation 1. In the figure, we use students' final high-school GPA as measure of prior performance. The high-school GPA is recoded to run from 1=worst to 4=best. In each panel, the first column uses the baseline sample. After the first column, the even-numbered columns use OLS regressions. The odd-numbered columns instrument a dummy for initial program take-up by the (random) treatment assignment variable. Columns (2) and (3) use students in lowest tercile of the distribution of high-school GPAs. Columns (4) and (5) use those in the middle tercile and Columns (6) and (7) those in the highest tercile. The terciles are differentially large because high-school GPA is not completely continuous. The regressions in Panel (a) use the number of credits for which students registered in the summer term 2020 as the dependent variable. The regressions in Panel (b) use the number of earned credits in the summer term as the dependent variable. The regressions in Panel (c) use students' average GPA (running from 1=worst to 4=best) among earned credits in the summer term as the dependent variable. The number of observations differs from Panels (a) and (b) since we have several students who do not earn any credits. We also drop 11 observations where we do not have information on students' high-school GPA. Standard errors are robust. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

B.4 More Heterogeneity by Gender

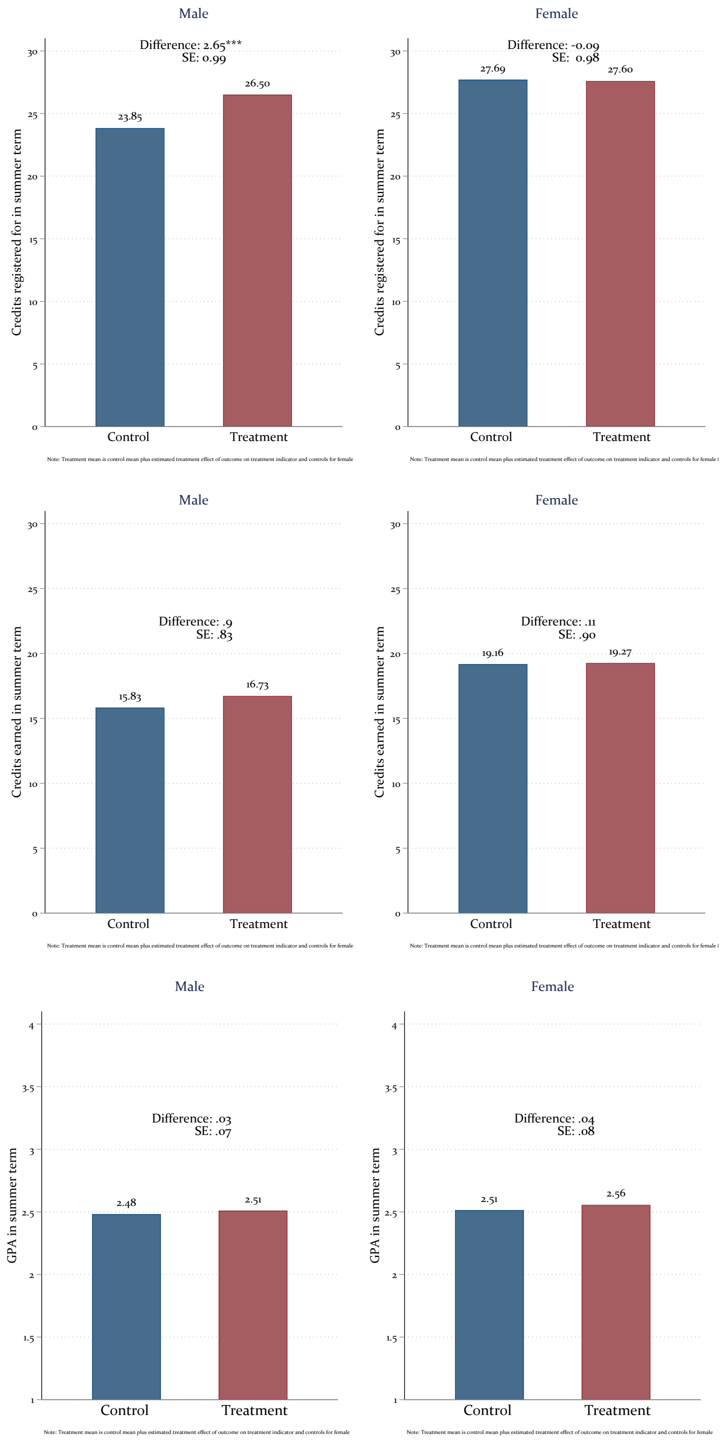


Figure B.4: Mean Differences by Mentee Gender

Note: This figure shows student outcomes by treatment status and gender. The top panel uses credits for which students registered in the summer term 2020 as outcome. The middle panel uses the credits earned in the summer term as outcome. The bottom panel uses average GPA (running from 1=worst to 4=best) among earned credits as outcome. In all panels, the control mean is calculated as the students' mean in the control group. Treatment effects, reported in the top center of each panel, are estimated using an OLS regression of the outcome on a treatment indicator, an indicator for students' gender, and students' credits earned in their first term. The treatment mean in the panel is calculated as the control mean plus the estimated treatment effect. Standard errors reported are robust.

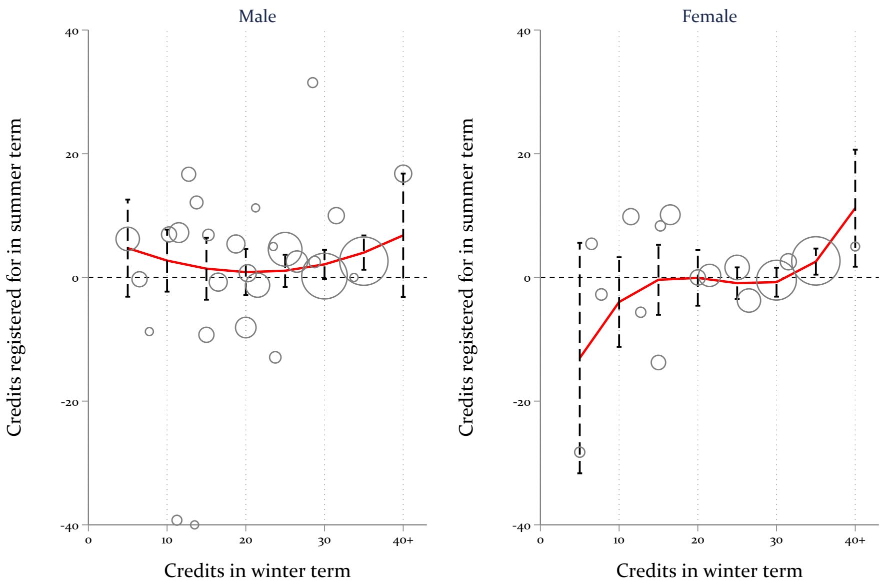


Figure B.5: Effects on Credits Registered for by Gender, by Credits Earned in Winter Term

Note: This figure shows how students' credits for which they registered in the summer term 2020 relate to students' prior performance as measured by students' credits earned in the winter term. The left panel shows the results for male students while the right panel shows the results for female students. The bubbles represent empirical differences between treatments, and the red solid lines indicate the treatment effects obtained from the model $y_i = \sum_{j=0}^{3} \beta_j \cdot (x_i)^j + \sum_{j=0}^{3} \gamma_j \cdot (x_i)^j \cdot T_i + u_i$, where y_i is the outcome of interest, x_i is our measure of prior performance, and T_i is an indicator for the treatment group. The spikes indicate 95% confidence bands (Huber-White standard errors). One student in the sample passed 45 credits in the winter term and is included in group "40+" for better visibility.

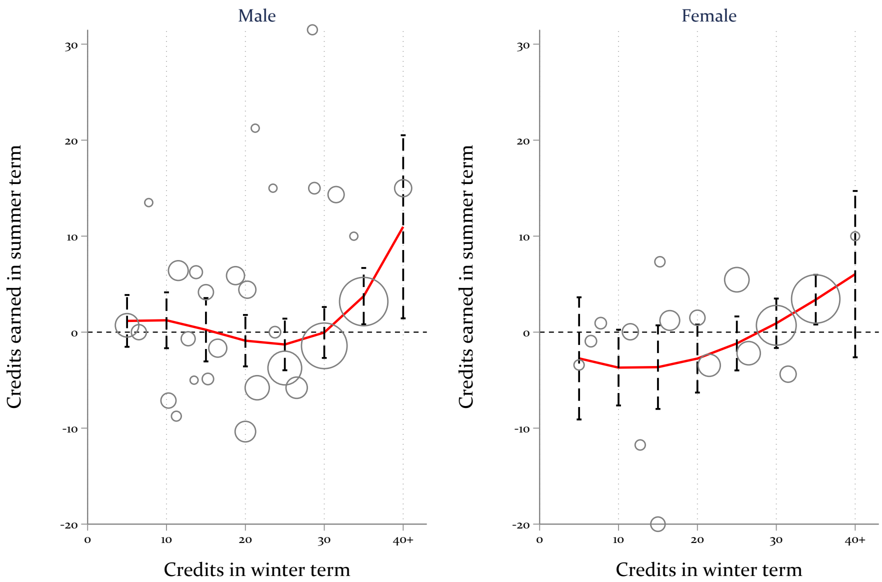


Figure B.6: Effects on Credits Earned by Gender and by Credits Earned in Winter Term

Note: This figure shows how students' credits earned in the summer term 2020 relate to students' prior performance as measured by students' credits earned in the winter term. The left panel shows the results for male students while the right panel shows the results for female students. The bubbles represent empirical differences between treatments, and the red solid lines indicate the treatment effects obtained from the model $y_i = \sum_{j=0}^{3} \beta_j \cdot (x_i)^j + \sum_{j=0}^{3} \gamma_j \cdot (x_i)^j \cdot T_i + u_i$, where y_i is the outcome of interest, x_i is our measure of prior performance, and T_i is an indicator for the treatment group. The spikes indicate 95% confidence bands (Huber-White standard errors). One student in the sample passed 45 credits in the winter term and is included in group "40+" for better visibility.

B.5 Endogenous Stratification

Table B.2: Endogenous Stratification

Panel A: Credits registered for			
Predicted Outcome Group:	Low	Middle	High
	(1)	(2)	(3)
Repeated split sample			
Coefficient	0.91	1.03	1.47
Std. Err.	1.45	0.94	0.90
Leave-one-out			
Coefficient	0.98	0.80	1.56
Std. Err.	1.59	1.22	1.01

Panel B: Credits earned			
	(1)	(2)	(3)
Repeated split sample			
Coefficient	-1.19	-0.15	2.76
Std. Err.	1.01	1.11	0.93
Leave-one-out			
Coefficient	-1.38	0.08	2.30
Std. Err.	1.06	1.19	0.98

Panel C: GPA			
	(1)	(2)	(3)
Repeated split sample			
Coefficient	0.13	-0.01	-0.00
Std. Err.	0.09	0.08	0.08
Leave-one-out			
Coefficient	0.14	-0.01	-0.04
Std. Err.	0.09	0.09	0.08

Note: This table shows impacts of peer mentoring on administrative student outcomes by students' predicted outcome group ("Group"), following the procedures outlined in Abadie et al. (2018) and using the Stata package `estrat` by Ferwerda (2014). We use students' gender, students' earned credits in the winter term, and students' high-school GPA as predictors. All regressions control for student gender and earned credits in the winter term. We use 100 RSS repetitions and 500 bootstrap repetitions, with 338 treated and 342 untreated observations in Panels A and B, since we do not have information on the high-school GPA of 11 students. The "low" group has 226 observations, the "middle" group 222 observations, and the "high" group 232 observations. In Panel C, we use 290 treated and 296 untreated observations since we cannot compute GPAs for students that do not pass any credits.

B.6 Treatment Effects by Students' Origin Region

Table B.3: Treatment Effects by Region of Origin

		Regional		Other	
Panel A: Credits registered for					
	ITT	ITT	IV	ITT	IV
	(1)	(2)	(3)	(4)	(5)
Treatment	0.19	1.37*	3.51*	-0.08	-0.17
	(1.57)	(0.78)	(1.98)	(1.57)	(3.37)
Treatment · regional	1.19				
	(1.76)				
Mean dep.	26.29	25.84	25.84	28.06	28.06
Obs.	679	542	542	137	137
Panel B: Credits earned					
	(1)	(2)	(3)	(4)	(5)
Treatment	1.23	0.37	0.96	0.85	1.86
	(1.51)	(0.66)	(1.67)	(1.57)	(3.38)
Treatment · regional	-0.84				
	(1.65)				
Mean dep.	17.66	16.69	16.69	21.48	21.48
Obs.	679	542	542	137	137
Panel C: GPA					
	(1)	(2)	(3)	(4)	(5)
Treatment	-0.01	0.04	0.10	0.01	0.01
	(0.10)	(0.06)	(0.14)	(0.10)	(0.20)
Treatment · regional	0.05				
	(0.12)				
Mean dep.	2.52	2.49	2.49	2.64	2.64
Obs.	585	464	464	121	121

Note: This table shows impacts of peer mentoring on administrative student outcomes by students' region of high-school graduation adapting equation 1. Students are labeled "regional" if they obtained their high-school graduation certificate in one of the three subregions of the region where the university is located. In each panel, the first column uses the baseline sample. After the first column, the even-numbered columns use OLS regressions. The odd-numbered columns instrument a dummy for initial program take-up by the (random) treatment assignment variable. Columns (2) and (3) use local students. Columns (4) and (5) use all other students. The regressions in Panel (a) use the number of credits for which students registered in the summer term 2020 as the dependent variable. The regressions in Panel (b) use the number of earned credits in the summer term as the dependent variable. The regressions in Panel (c) use students' average GPA (running from 1=worst to 4=best) among earned credits in the summer term as the dependent variable. The number of observations differs from Panels (a) and (b) since we have several students who do not earn any credits. Standard errors are robust. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

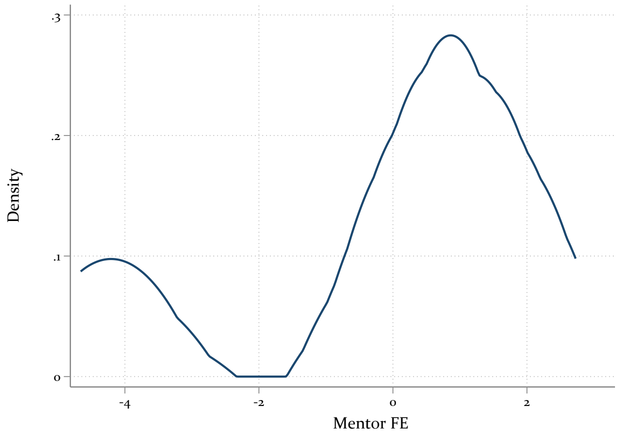
B.7 Mentor "Value-Added"

In this subsection, we provide results from a sort of "value-added" analysis where we regress students' outcomes in the summer term on their performance in the winter term, their observable characteristics, and a mentor dummy. We label the estimated mentor fixed effect in each outcome dimension the mentor's "value-added". Specifically, we estimate the following equation:

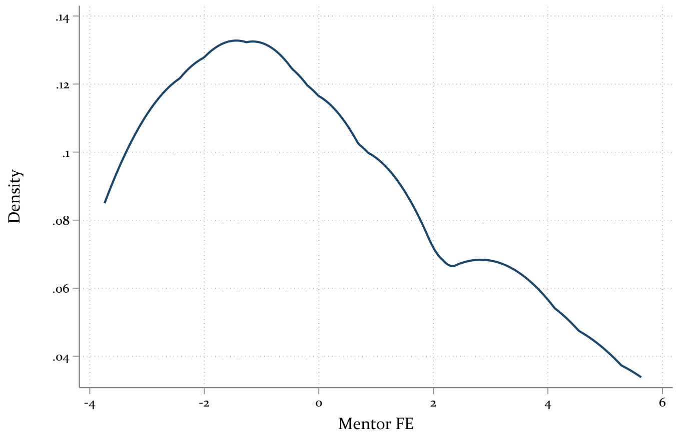
$$y_i = \alpha + x_i\beta + \mu_j + \epsilon_i \tag{2}$$

where y_i is the respective outcome of student i in the summer term, X_i is a vector of observable characteristics including past performance (credits earned in the winter term and high-school GPA), students' gender and age, a dummy whether students are from the region where the university is located, whether students obtained their university entrance qualification abroad, whether they are part-time students, and whether the are enrolled at university for the first time. μ_j is the fixed effect of mentor j, which we interpret as mentors' value-added. Standard errors are robust.

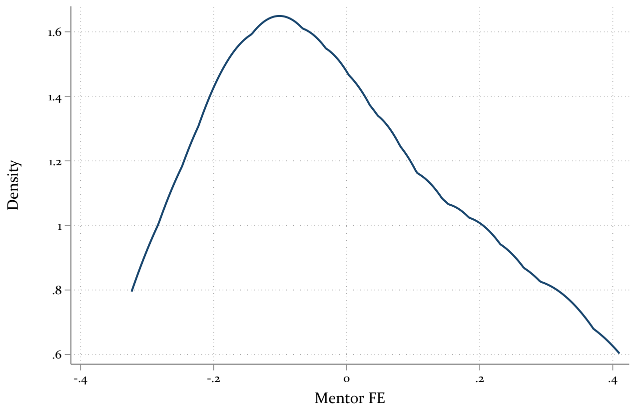
Figure B.7 shows the results. We find that mentors differ substantially in their value-added. The value-added estimates on earned credits range from -3 credits to 7 additional credits conditional on students' observables. We also test whether the value-added estimates are correlated across different outcomes measures in Table B.4. While mentors' value-added on credits earned and registered for are strongly correlated (ρ=0.56,$p < 0.05$), both measures are not substantially correlated with value-added on GPA. We again caution that each mentor only advises up to 10 mentees, thus leading to substantial noise in the value-added estimates. However, the mentors were randomly assigned to mentees, such that we do not have the problem of endogenous sorting of students to mentors common in the literature on teacher value-added. We also only have 15 mentors, such that we cannot credibly identify the sources of mentors' performance differences.



(a) Credits registered for



(b) Credits earned



(c) GPA

Figure B.7: Kernel Density Plots of Mentor Value-Added

Note: This figure shows Kernel density plots of mentor value-added, μ_j, in Equation 2. Panel (a) uses the number of registered credits in the summer term 2020 as outcome measure. Panel (b) uses earned credits in the same term as outcome measure. Panel (c) uses GPA among earned credits (running from 1=worst to 4=best) as outcome measure.

	Mentor value-added (earned credits)
Mentor value-added (credits registered for)	0.525**
	(0.044)
Mentor value-added (GPA)	-0.060
	(0.832)

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.4: Pairwise Correlations of Mentor Value-Added

B.8 Treatment Effects by Subject

Table B.5: Treatment Effects by Module

Panel A: ITT Estimates							
Dep. Var.:	# passed	Passed in ST					# passed
	Core ST	Sales	Fin. Statements	Macro	Micro	Fin. Maths.	Core WT
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.08	-0.04	-0.02	0.05*	0.06*	0.03	-0.05
	(0.09)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)
Mean dep.	2.34	0.68	0.47	0.39	0.42	0.37	
Obs.	691	691	691	691	691	691	691
Panel B: IV Estimates							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.19	-0.09	-0.06	0.13*	0.14*	0.06	-0.12
	(0.22)	(0.07)	(0.08)	(0.07)	(0.07)	(0.09)	(0.09)
Mean dep.	2.34	0.68	0.47	0.39	0.42	0.37	
Obs.	691	691	691	691	691	691	691

Note: This table shows impacts of peer mentoring on the likelihood of passing specific core courses, adapting equation 1. Panel (A) uses OLS regressions. Panel (B) instruments a dummy for initial program take-up by the (random) treatment assignment variable. Column (1) uses the sum of passed summer term core courses in the summer term 2020 as the dependent variable. The core courses in the summer term are Sales, Financial Statements, Macroeconomics, Microeconomics, Financial Mathematics, and a language course (not shown). Columns (2) through (6) use separate dummy for each core course as dependent variable. Column (7) uses the sum of passed courses in the summer term that were core courses in the winter term as dependent variable. Standard errors are robust. * p<0.10, ** p<0.05, *** p<0.01

C Additional Survey Information and Evidence

C.1 Survey Questions

In this table, we show the exact questions that we asked students in the survey. The survey was conducted in German, which is the official language of the program. In addition to these questions, we asked students to list the exams they intend to sit, module by module.

Label	**Question**
Motivation	I was able to motivate myself well during the virtual summer semester.
Continuous studying	I was able to cope well with the challenge of continuously studying for courses during the virtual summer semester.
Timely exam prep.	In the virtual summer semester, I started my exam preparation on time.
Clear contact person	I was always able to find a suitable contact person for questions and problems concerning my studies.
Sufficient effort	Measured against my goals for this semester, my effort to study during the lecture period was sufficient.
Student services	I am satisfied with the individual services offered by the School of Business, Economics and Society during the virtual summer semester.
Communication	Overall, I am satisfied with the way the School of Business, Economics and Society communicated during the virtual summer semester.
Cares for my success	I feel that the people in charge at the School of Business, Economics and Society care for my academic success.
Takes my concerns seriously	I feel that my individual concerns and problems as a student are taken seriously at the School of Business, Economics and Society.
Online content	I am satisfied with how the online teaching was implemented content-wise in the virtual summer semester.
Technical implementation	I am satisfied with how the online teaching was technically implemented in the virtual summer semester.
Contact to other students	During the virtual summer semester, I regularly had contact to other students from my semester to discuss study matters.
Can work in principle	Based on my experiences in the virtual summer semester, I believe that online teaching at the university can work well in principle.
Should play large role	Based on my experiences in the virtual summer semester, I believe that online teaching should play an important role at university in the future.
Prob. timely graduation	I estimate the probability that I will complete my studies within the designated period of study (six semesters) at [ ] percent.

Table C.1: Survey Questions

C.2 Regression Results

Table C.2: Treatment Effects on Assessment of Own Motivation and Study Effort

Dep. Var.:	Motivation		Continuous studying		Timely exam prep.		Sufficient effort	
	ITT	IV	ITT	IV	ITT	IV	ITT	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.22*	0.46*	0.19*	0.40*	0.11	0.23	0.20**	0.42**
	(0.12)	(0.25)	(0.11)	(0.23)	(0.11)	(0.22)	(0.10)	(0.21)
Mean dep.	2.71	2.71	2.93	2.93	2.99	2.99	3.18	3.18
Obs.	404	404	404	404	404	404	404	404

Note: This table shows impacts of peer mentoring on on survey outcomes, adapting equation 1. The odd-numbered columns use OLS. The even-numbered columns use (random) treatment assignment variable as an instrument for initial program take-up. All dependent variables are measured on a five-point Likert scale where higher outcomes indicated more agreement with the question. The questions underlying the dependent variables are: (Columns 1 and 2); (Columns 3 and 4); Columns (5 and (6); and (Columns 7 and 8). Standard errors are robust. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table C.3: Treatment Effects on Assessment of Department Services

	Department services						Department relations			
Dep. Var.:	Support service		Communication		Clear contact person		Cares about my success		Takes my concerns seriously	
	ITT	IV	ITT	IV	ITT	IV	ITT	IV	ITT	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment	0.01	0.03	-0.05	-0.09	0.07	0.15	0.11	0.22	0.02	0.04
	(0.09)	(0.18)	(0.10)	(0.20)	(0.10)	(0.22)	(0.10)	(0.21)	(0.10)	(0.20)
Mean dep.	3.51	3.51	3.56	3.56	3.23	3.23	3.16	3.16	3.44	3.44
Obs.	404	404	404	404	404	404	404	404	404	404

Note: This table shows impacts of peer mentoring on on survey outcomes, adapting equation 1. The odd-numbered columns use OLS. The even-numbered columns use (random) treatment assignment variable as an instrument for initial program take-up. All dependent variables are measured on a five-point Likert scale where higher outcomes indicated more agreement with the question. The questions underlying the dependent variables are: (Columns 1 and 2); (Columns 3 and 4); Columns (5 and (6); and (Columns 7 and 8). Standard errors are robust. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table C.4: Treatment Effects on General Views on Online Teaching

	Summer Term 2020						Online teaching in general			
Dep. Var.:	Online content		Technical implementation		Contact to other students		Can work in principle		Should play larger role	
	ITT	IV	ITT	IV	ITT	IV	ITT	IV	ITT	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment	-0.06	-0.12	0.03	0.06	0.03	0.07	0.06	0.13	0.10	0.20
	(0.10)	(0.20)	(0.09)	(0.20)	(0.14)	(0.28)	(0.11)	(0.24)	(0.14)	(0.29)
Mean dep.	3.52	3.52	3.73	3.73	2.74	2.74	3.60	3.60	3.40	3.40
Obs.	404	404	404	404	404	404	404	404	404	404

Note: This table shows impacts of peer mentoring on on survey outcomes, adapting equation 1. The odd-numbered columns use OLS. The even-numbered columns use (random) treatment assignment variable as an instrument for initial program take-up. All dependent variables are measured on a five-point Likert scale where higher outcomes indicated more agreement with the question. The questions underlying the dependent variables are: (Columns 1 and 2); (Columns 3 and 4); Columns (5 and (6); and (Columns 7 and 8). Standard errors are robust. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

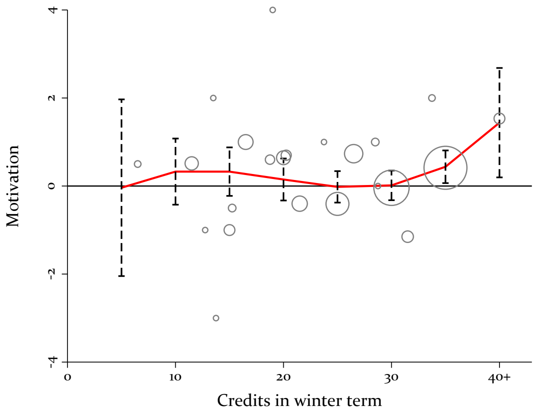
C.3 Average Standardized Effects

Table C.5: Average Standardized Effects on Survey Outcomes

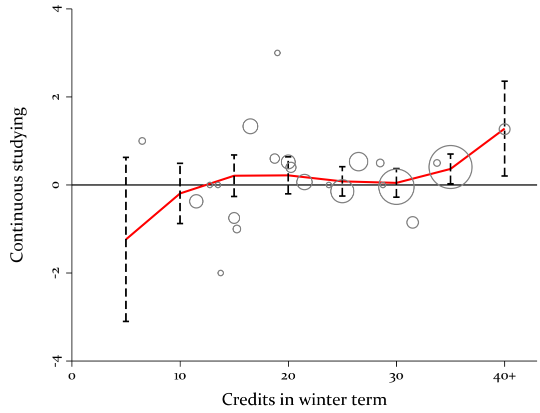
	Assessment of..		
	Study behavior	Department services	Online teaching
	(1)	(2)	(3)
Avg. effect	0.16**	0.03	0.02
	(0.08)	(0.08)	(0.07)
Obs.	404	404	404

Note: This table shows average standardized effects on survey outcomes by broad group using the methodology of Kling et al. (2004) and Clingingsmith et al. (2009). Standard errors are robust. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

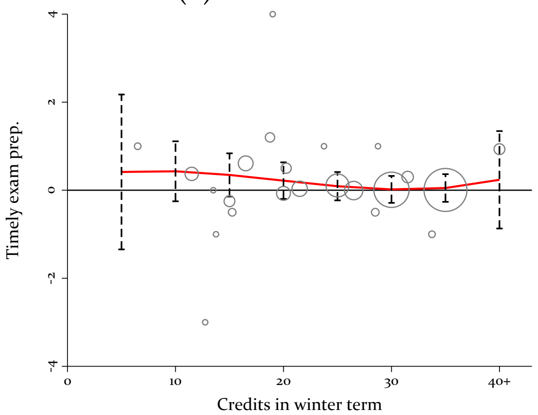
C.4 Heterogeneity of Treatment Effects on Survey Outcomes by Credits Earned in Winter Term



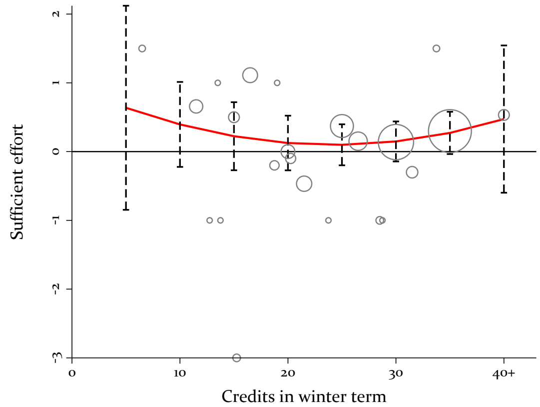
(a) Motivation



(b) Continuous studying



(c) Timely exam prep.



(d) Sufficient effort

Figure C.1: Treatment Effects on Survey Outcomes by Credits Earned in Winter Term

Note: This figure shows how students' survey responses relate to students' prior performance as measured by students' registered credits in the winter term, which was students' first term at university. Panels (a) to (d) display heterogeneous treatment effects (relative to the control group) on students' responses to questions about their motivation during their summer term, their self-assessed ability to study continuously in the summer term, their self-assessment of whether they prepared for exams timely, and their assessment of whether they studied enough to meet their goals in the summer term, respectively. All responses are measured on a five-point Likert scale where higher values signal more agreement with the question or statement. The bubbles represent empirical differences between treatments, and the red solid lines indicate the treatment effects obtained from the model $y_i = \sum_{j=0}^{3} \beta_j \cdot (x_i)^j + \sum_{j=0}^{3} \gamma_j \cdot (x_i)^j \cdot T_i + u_i$, where y_i is the outcome of interest, x_i is our measure of prior performance, and T_i is an indicator for the treatment group. The spikes indicate 95% confidence bands (Huber-White standard errors). One student in the sample passed 45 credits in the winter term and is included in group "40+" for better visibility.