CESifo

AREA CONFERENCES 2021

Economics of Education

Munich, 3–4 September 2021

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

David Hardt , Markus Nagler, and Johannes Rincke

CESifo

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

David Hardt[†], Markus Nagler[‡], Johannes Rincke[§]

August 10, 2021

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.

*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, Christine Binzel, Ben Castleman, Fabian Dehos, Ulrich Glogowsky, Joshua Goodman, Philipp Lergetporer, Claus Schnabel, Martin Watzinger, Kathrin Wernsdorf, Martin West and participants at various research seminars and conferences 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.

[†]University of Erlangen-Nuremberg; david.hardt@fau.de

[‡]University of Erlangen-Nuremberg, CESifo, and LASER; markus.nagler@fau.de

[§]Corresponding author; University of Erlangen-Nuremberg and CESifo; johannes.rincke@fau.de

1 Introduction

Online delivery of tertiary education is on the rise. The COVID-19 pandemic has forced virtually all education institutions to switch to online teaching. The literature has generally found online teaching to be inferior to classroom-based teaching (e.g., Figlio et al., 2013; Bettinger et al., 2017). Switching to online teaching may thus aggravate the problem that many students struggle to complete their studies successfully (Weiss et al., 2019). Accordingly, students expect and experience negative consequences of the COVID-19-induced shift to online teaching (Aucejo et al., 2020; Bird et al., 2020; Kofoed et al., 2021). This may be due to problems of disorganization among students in online teaching, as argued for massive open online courses ("MOOCs"; e.g. Banerjee and Duflo, 2014; McPherson and Bacow, 2015; Patterson, 2018). One way to improve outcomes of online education could therefore be to assist students through online peer mentoring. Evidence on the effectiveness of such programs is scarce for online teaching, where they may be particularly helpful.

In this paper, we report results of a randomized trial studying the effects of peer mentoring at a German university that, due to the COVID-19 pandemic, switched to online teaching for the summer term 2020. Our sample comprises of 691 second term students from the core undergraduate program at the university's *School of Business and Economics*. To assess the effectiveness of the program, we combine registry data with survey data on study behavior and motivation. Our paper presents the first evidence on the role of remote peer mentoring in online higher education.

The mentoring program focused on students' general study skills, such as self-organization and study techniques. Mentors and mentees met one-on-one online. The program consisted of five meetings of around 40 minutes each that took place around every two weeks. In each meeting, mentors would discuss specific topics, such as mentees' weekly study schedules, using materials provided by us, as well as follow-on discussions on prior topics. Importantly, we instructed mentors not to discuss any coursework with mentees. As mentors, we hired students from a more advanced term in the same study program. Thus, this kind of mentoring could be scaled up easily and at modest cost.

Our setting is common for public universities across the developed world. Each fall, students enroll in the three-year bachelor's program *Economics and Business Studies*. In each of the first two semesters, students are to pass six courses each worth five credits. Since the second term includes more rigorous courses relative to the first semester, many students struggle in this term.[1] A key advantage of our setting is that

[1] Administrative data from the year 2018/19 shows that even in regular times, many students underperform relative to the suggested curriculum: after the first semester, only 59 percent of enrolled students have completed courses worth at least 30 credits.

the summer term 2020 was conducted *entirely* online because the German academic summer term starts and ends later (April to July) than is common internationally.

Our main results are as follows. First, the mentoring program improved students' motivation and study behavior. Treated students report higher overall motivation, are more likely to report having studied throughout the term, and are more likely to state they provided enough effort to reach their term goals. Second, while these effects translate into more exam registrations, the average effect on earned credits is small and insignificant. Similarly, the students' GPA is unaffected. Third, these results mask a heterogeneity that contrasts with common expectations on potential impacts of peer mentoring. For instance, we observe a positive effect on students who previously performed well, with no effects on other students. In addition, male students benefit more from the program, if anything. These results somewhat contrast prior research suggesting that weaker students struggle most in online learning (e.g., Figlio et al., 2013; Bettinger et al., 2017) and that female students tend to benefit more from mentoring (e.g., Angrist et al., 2009; Rodriguez-Planas, 2012).

We contribute to research on the online education production function. This literature has found online teaching to be less effective than classroom-based teaching Figlio et al. (2013); Bettinger et al. (2017), likely due to problems of disorganization among students in online teaching (e.g. Banerjee and Duflo, 2014).[2] Research on specific aspects of the online education production function is scarce. Closest to our work, Oreopoulos et al. (forthcoming) show that assigning online and regular students to an online planning module to construct weekly study schedules and reminders or loose mentoring via text messages does not affect students' outcomes. We focus on a more comprehensive and intensive mentoring program with regular contact and guidance, which has been shown to matter (e.g., Oreopoulos and Petronijevic, 2018). We thus contribute by providing the first evidence on the effectiveness of peer mentoring programs for online higher education.

We also contribute to the experimental literature on mentoring interventions in higher education.[3] Closest to our work, Oreopoulos and Petronijevic (2018) experimentally study different coaching methods. They find that close one-on-one peer mentoring programs are effective in raising student outcomes. Oreopoulos and Petronijevic (2019) show evidence from several nudging interventions that did not

[2]In line with this, Patterson (2018) experimentally studies commitment devices, alerts, and distraction blocking tools in a MOOC and finds positive effects for treated students. Delivery-side frictions such as lack of experience in online teaching may also be important (e.g., Orlov et al., 2020).

[3]Lavecchia et al. (2016) provide a recent review of behavioral interventions in education production. For other higher-education interventions, see, e.g. Bettinger et al. (2012); Himmler et al. (2019); Clark et al. (forthcoming).

shift students' academic outcomes, but improved their motivation. Angrist et al. (2009) test the impact of academic support and financial incentives on students' GPA, finding performance increases among female students. However, they find no effects for academic support without financial incentives. Our program is targeted more towards individual mentor-mentee interactions, is more structured, and it takes place in an online environment where mentoring may be more important. We thus contribute by providing the first evidence on the effectiveness of close (peer) mentoring in online education and by extending the small experimental literature on mentoring effects in higher education.[4]

Finally, we contribute to research on education responses to the COVID-19 pandemic, most of which has focused on primary or secondary education (e.g., Angrist et al., 2020; Bacher-Hicks et al., 2020; Grewenig et al., 2020). The closest paper is Carlana and Ferrara (2020), who experimentally assigned Italian middle school students an online mentor during the pandemic and report positive effects on performance and well-being. We contribute by studying the effectiveness of online mentoring in higher education. Despite the universal shift towards online teaching in higher education due to the pandemic, evidence on improving the effectiveness of online teaching in this context remains scarce. This is despite early evidence suggesting that the shift led to worse outcomes (Bird et al., 2020; Kofoed et al., 2021).

2 Experimental Setting and Design

2.1 Experimental Setting

Our setting is typical of public universities in the Western world. The undergraduate study program *Economics and Business Studies* at the intervention university requires students to collect 180 credits to graduate, which is expected after three years. The study plan assigns courses worth 30 credits to each semester. Administrative data show that large shares of students do not complete 30 credits per semester, delaying their graduation. Survey data collected from earlier cohorts of students suggests that most students do not work full-time even when summing up hours studied and hours worked to earn income. The salient study plan and target of achieving 30 credits per term, the fact that most students register for exams worth these credits, and the

[4]Bettinger and Baker (2014) show that a (professional) student coaching service focusing on aligning long-term goals and self-organization and providing study skills increased university retention. Castleman and Page (2015) provide evidence of an effective text messaging mentoring for high-school graduates. CUNY's ASAP program combines several interventions from financial support to tutoring and seems highly effective (Scrivener et al., 2015; Sommo et al., 2018; Weiss et al., 2019).

fact that students do not seem to work enough to pass these exams suggests that many students have problems in self-organizing and/or studying efficiently. Given prior findings on such problems in online education, we expected these issues to be 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 university were conducted online. To this end, the university relied on *Zoom*, an online video tool used widely during the pandemic at universities around the globe. A key advantage of our setting is that the summer term 2020 was conducted entirely online because the German academic year starts and ends later than common internationally.[5] It is therefore cleaner than would be possible in other settings since there are no spillovers from in-person to online teaching.

2.2 The Mentoring Program

In the first term week, students in the treatment group were informed via e-mail about the launch of a mentoring program for students in the second semester. They were invited to register for the program through a webpage.[6]

The program focused on self-organization and on making mentees aware of potential problems of studying online. We designed it to involve five one-on-one online meetings between mentors and mentees, taking place around every two weeks. The average length of meetings reported by the mentors was around 40 minutes. For each meeting, we provided mentors with some information. Because our sample is rather small, we combined several aspects of mentoring that the prior literature has suggested to be effective into one program.

The first meeting focused on mentees' expectations of their term performance and contrasted these expectations with the average performance of previous cohorts to target student overconfidence (Lavecchia et al., 2016). Mentors were also provided advice on self-organization when working from home, targeting student disorganization in online environments (e.g., Banerjee and Duflo, 2014). In the second meeting, mentors and mentees formulated specific goals for the mentee. This included study goals (weekly study schedule, see Figures A.1 and A.2 in the Appendix), courses to be taken, and performance-based goals (credits earned), based on research on the effectiveness of goal-setting (e.g., Clark et al., forthcoming). The third meeting

[5]Teaching started on April 20th and the exam period started on July 20th, 2020.

[6]The page asked for the students' consent to use their personal information for research in anonymized form and for their consent to pass along names and e-mail addresses to mentors. Treatment group students who did not register for the program within two days received reminders.

focused on exam preparation (timing of exams, implications for mentee's preparation), targeting students' alignment of long-term goals and short-term behavior (e.g., Bettinger and Baker, 2014).

The fourth meeting focused on studying effectively. This included the presentation of a simplified four-stage learning model (see Figure A.3 in the Appendix) and how to implement these learning strategies in practice, targeting students' study skills (e.g., Angrist et al., 2009). The final meeting focused on the mentee's exam preparation, including a time schedule providing guidance on how to specifically prepare for exams. This targeted students' underestimation of the time required to complete a task (e.g., Oreopoulos et al., forthcoming). In all meetings, mentors and mentees additionally discussed issues the mentee was currently facing, similar to general counseling services in other settings (e.g., Rodriguez-Planas, 2012). We instructed mentors to ensure that the information was only provided to mentees and not to other students.

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

2.3 Recruitment and Training of Mentors

We hired 15 peer mentors, with each mentor handling ten mentees at most. The mentoring program's capacity was therefore 150 students.[7] All mentors were students who successfully completed the first year and were enrolled in the fourth semester of the study program during the summer term. They had good GPAs and high-school GPAs and more likely worked in student jobs next to their studies. Among all applicants, we selected those we felt would be the most able mentors. Eight of the mentors were female and seven were male.

All mentors took part in an online kick-off meeting where we explained the purpose and the structure of the program and laid out the planned sequence and contents of the mentoring sessions. They were not informed that the program was experimentally evaluated, but were informed that the program's capacity was limited and that a random subset of students in the second term was invited to participate. They subsequently took part in a training by professional coaches. The training focused on communication skills and took about five hours. Three weeks after program start,

[7]The program could be scaled up easily and at low cost. Including one additional mentee for a three-month period would cost about €60. Mentors were employed for three months, with work contracts on four hours per week and monthly net pay of about €160. Employer wage costs were about €200 per month and mentor.

the mentors took part in a short supervision meeting with the coaches. In addition, we sent regular e-mails to the mentors (before each of the five meetings) and answered questions.

2.4 Sampling and Assignment of Students to Mentors

About 850 students enrolled for the study program *Economics and Business Studies* in the fall of 2019. We excluded students who dropped out after the first semester, who were not formally in their second semester in the summer term 2020 (e.g., because of having been enrolled with some credits at another university before), and who completed less than a full course (5 credits) in the first term. This leaves us with 694 students. We randomly assigned half of the students to treatment using a stratified randomization scheme with gender and earned credits in the first term as strata variables. After the intervention ended, we had to drop another three students from the sample who got credited for second-term courses earned elsewhere.[8] Our 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 completed up to 30 credits in their first term (369 students). We then successively invited three further groups of students sampled into treatment according to their credits earned in the first semester, until all 344 students sampled into treatment were invited. 142 of these students signed up for the mentoring. We randomly assigned participating students to mentors. To achieve a balanced gender mix of mentee-mentor pairs, we used the mentees' gender as a strata variable in the assignment. Among registered mentees, about 54 percent were female. Thus, the number of pairs in each of the mentee-mentor gender combinations was similar.

3 Data and Empirical Strategy

3.1 Data

Survey Data

After the fifth round of mentee-mentor meetings, we invited all 691 students in the experimental sample to an online survey. It was conducted on an existing platform at the department that is regularly used to survey students. Students who completed the survey, which lasted around ten minutes, received a payoff of €8.00. The survey

[8]Students are free when to hand in certificates on credits earned elsewhere, delaying this information in the administrative data.

elicited the students' assessment of their study effort, their satisfaction with the department's effort to support online learning during the term, views on online teaching generally, and beliefs about one's academic achievement. The full set of questions is shown in Appendix B.1. We use all survey responses submitted until the beginning of the examination period to avoid spillovers from exams to the survey. 404 students (58.5% of the sample) participated.

Administrative Data

We collected registry data from the university in mid October 2020 to measure outcomes related to academic achievement in 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 to measure attempted examinations, interpreted as student effort. While students were not materially harmed by not passing exams in this summer term, registering may be costly since it may generate opportunity costs. Our primary outcome is, second, credits earned in the summer term. This measures most directly the students' academic achievement during the intervention term. Note, however, that this might be a slow-moving variable since study effort has cumulative gains over time. Following Angrist et al. (2009), we did not exclude students who withdrew from the sample. These students were coded as having zero attempted and earned credits. 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 grade).[9] Given that we expect (and find) impacts of the treatment on the 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 achievement is students' GPA. The reason for this difference is that in the German system, students are typically free to choose the timing of taking their courses even when a core curriculum is suggested. In addition, many students do not attempt to complete the curriculum in the suggested time period, making the extensive margin of how many courses to take more relevant than in the U.S.

The exams took place after the end of the teaching period between end of July and September 2020. In addition, the university provided us with background information on individual students. The characteristics include information on enrollment, gender, age, type of high school completed, and information on high-school GPA (coded from 1 as the worst to 4 as the best grade).

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

Table 1: Summary Statistics by Treatment Status

	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.

3.2 Balancing Checks and Take-Up

Balancing Checks

Table 1 reports differences in means and standardized differences in students' characteristics. The characteristics 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 being in their first year at university, and a dummy for full-time students.[10] As can be seen from Table 1, the treatment and control groups were well balanced across all characteristics.

To assess the quality of our survey data, we repeat the balancing checks using our survey respondents. We also study selection into survey participation by mean-comparison tests between survey participants and non-participants. Table B.2

[10]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 studies are integrated into a vocational training program.

Table 2: Take-Up

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

in the Online Appendix shows that students who participated in the survey differ slightly from students who did not participate. Participants are somewhat younger, more female, have better high-school GPA, have earned more credits in the winter term, and are more likely part-time students. Importantly, the likelihood survey completion is unrelated to treatment assignment. Within the sample of participants, treatment and control group are balanced across all characteristics.

Take-Up

142 students signed up for the program. Table A.1 in Appendix A.2 shows that these students are slightly older, more likely female, and more likely to have a foreign university entrance exam than treatment group students who did not sign up. The differences are small, however. Students who registered for the program could drop out at any time with no penalty. Table 2 shows the program take-up and the first stage of our IV estimations. 41 percent of treatment group students signed up for the program (Column 1). Some students drop out before the first meeting, leaving 37 percent of those invited to taking at least one meeting (Column 2), whereas 32% take all five meetings (Column 3). The final two columns show that female students are more likely to sign up conditional on receiving an invitation, as in Angrist et al. (2009).[11]

[11] Section A.3 in the Appendix shows that female and male mentors differ slightly in their mentoring. They do not seem to be differentially effective, though.

3.3 Estimation

To evaluate the effects of the peer mentoring program, we estimate the equation

$$y_i = \alpha + \beta Treatment_i + \gamma_1 Female_i + \gamma_2 CreditsWT_i + \epsilon_i, \tag{1}$$

where y_i is the outcome of student i, $Treatment_i$ is an indicator for (random) treatment assignment, $Female_i$ is a dummy for female students, 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. Each of the outcomes is thus regressed on the treatment indicator and the strata variables. We report robust standard errors. Since not all treatment group students registered for the mentoring services, we additionally run IV regressions using the treatment assignment as an instrument for actual take-up. The main variable for measuring 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.

For several reasons, we considered it likely that the effects would be heterogeneous. First, prior evidence on online education shows more negative effects for weaker students (e.g., Figlio et al., 2013; Bettinger et al., 2017). We thus expected heterogeneous effects by credits earned in the first term.[12] Second, male students suffer more from online relative to in-person teaching (e.g., Figlio et al., 2013; Xu and Jaggars, 2014). However, take-up rates in mentoring programs seem 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 students, the relative effect of having been offered a mentor on outcomes, and the relative effect of mentoring on outcomes conditional on take-up, was unclear. We study treatment effect heterogeneities by including an interaction between the variable capturing the dimension of heterogeneity and the treatment indicator, joint with the variable capturing the dimension itself.

We investigated additional heterogeneities that we described as less central (and likely not to be reported) in the pre-analysis plan. First, whether the effects of mentoring are larger when mentored by female than by male mentors as well as gender interactions (e.g., Dee, 2005, 2007; Hoffmann and Oreopoulos, 2009). We study this in Table A.3 in the Appendix, but find no such results. Second, the pre-analysis plan also specified that we would test if the effect on students enrolled at university

[12]In addition, there is a positive baseline correlation between students' high school GPA and their university performance. In Appendix C.2, we therefore also show estimates using mentees' high-school GPA as the dimension of effect heterogeneity; results are similar.

for the first time differs from students who had been enrolled before. Again, this is not the case (not reported).

4 Results

4.1 Effects on Study Behavior and Motivation

We first study the effects of the mentoring program on self-reported study behavior and motivation and contrast those with effects on the perception of department services and online teaching more generally. Figure 1 shows results from OLS and IV estimations instrumenting take-up by treatment assignment. We show the treatment effects along with 90% confidence intervals. All dependent variables are responses on a five-point Likert scale, with higher values indicating higher agreement with the statement.[13]

Panel (a) shows treatment effects on students' assessment of their motivation and study behavior in the summer term. The mentoring program specifically targeted these outcomes. The first two rows show positive impacts on students' motivation. The estimated effect in the IV estimation amounts to around 18% of the control group mean. The next two rows show significant effects on students' response to whether they managed to study continuously throughout the term. The subsequent two rows show smaller effects on students' response to 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 B.6. This part of the survey shows an average standardized effect of around 0.16 standard deviations (p-value = 0.048).[14]

Panel (b) shows that the treatment did not shift views on departmental services generally, an aspect that the mentoring program was not directly concerned with. The

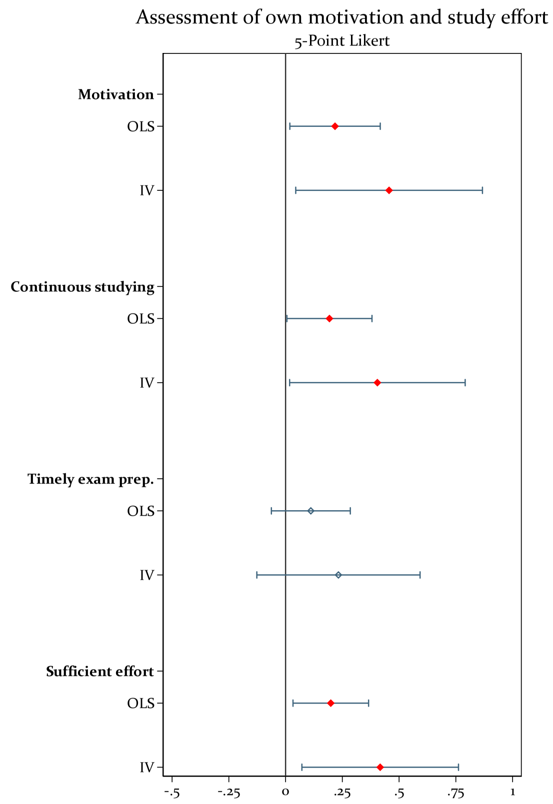
[13] All corresponding tables are in Online Appendix B. The statements are shown in Online Appendix B.1.

[14] In unreported analyses, we also investigate whether our effects are robust to adjustments to multiple hypothesis testing. First, we follow Kling et al. (2007) and estimate impacts on a Z-score index of all questions in this block. The ITT is 0.16 SD (p-value = 0.05) and the TOT shows an effect of 0.34 SD (p-value = 0.05). Second, we use randomization inference procedures (Heß, 2017). Our standard errors are unaffected. Third, we use the procedure by Barsbai et al. (2020) that corrects for multiple hypothesis testing. Here, the p-values are larger, at around .15 for the questions that are significant above. While these results are thus a bit noisy, they are overall robust.

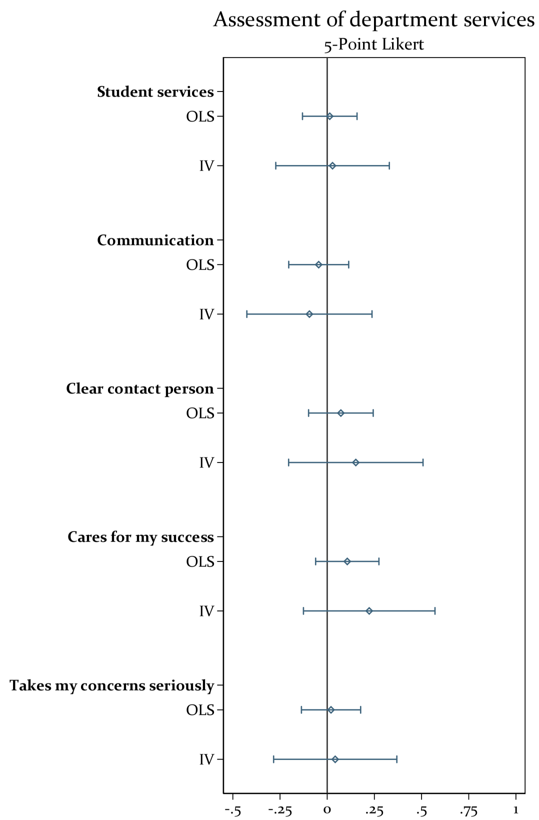
items include student services, communication by the department, whether there is a clear departmental contact person, and students' views on whether the department cares for their success or takes their concerns seriously. The most pronounced effect is for students' feeling whether the department cares for their success, with point estimates of 7% relative to the control group mean; this effect is however insignificant. Panel (c) reports results on students' general views on online teaching. The items include students' satisfaction with the departments' online teaching content and implementation in the summer term. We also asked students whether they feel online teaching can work in principle and whether it should play a large role in the future. Both effects are insignificant. We additionally analyze the students' response to the question whether they frequently interacted with other students. Here, the null result is interesting, since it shows that the program did not merely substitute for interactions among students.[15]

[15]We additionally elicited students' expectations of the likelihood of completing their studies in time and their planned credits. The results are noisy and show no difference between treatment and control group (not shown).

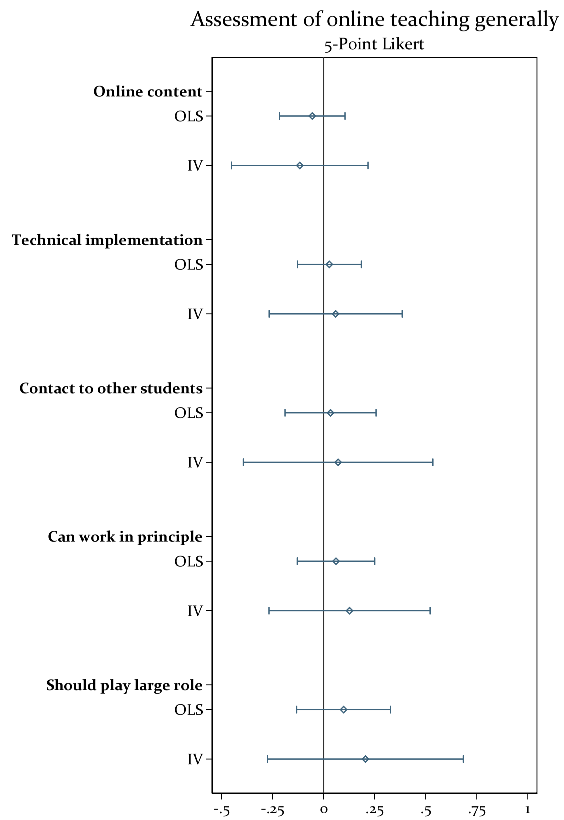
Figure 1: Impacts of Mentoring on Survey Outcomes



(a) Assessment of study behavior



(b) Assessment of department services



(c) Assessment of online teaching

Note: This figure shows impacts of peer mentoring on survey responses adapting Equation 1. For each question, the first row uses OLS regressions, estimating intent-to-treat effects (labeled "OLS"). The second row uses (random) treatment assignment as an instrument for initial program take-up, estimating treatment-on-the-treated effects (labeled "IV"). For the full set of survey questions, please see Online Appendix B.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 B.2. Standard errors are robust.

Overall, our results suggest that the peer mentoring program improved students' motivation and study behavior, hence working as intended. It is also reassuring that we do not see effects on survey items which are unrelated to peer mentoring. We now investigate whether these effects translated into improved academic outcomes.

4.2 Average Impacts on Primary Outcomes

Table 3 shows differences between treatment and control group for academic outcomes. The odd-numbered columns show OLS/ITT estimates, the even-numbered columns show corresponding IV/TOT estimates where we use treatment offer as an instrumental variable for program sign-up. Column (1) shows the impacts on credits registered for. Students who received a treatment offer register for around 1.4 more credits than students who did not. 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 (3) shows that students with a treatment offer earn around 0.5 more credits than control students. Registered students earn around 1.3 credits more (Column 4), which implies that students pass around 40% of the additional credits for which they register. These results are statistically insignificant, however. Finally, Columns (5) and (6) show that students' GPA is unaffected, indicating that the (modest) average increase in attempted credits did not come at the expense of worse average grades.

Overall, Table 3 suggests that the effects of the mentoring program on study behavior and motivation did translate into more exam registrations, but the impacts on performance (in terms of credits earned) are too noisy to rule out either zero effects or very large effects. To obtain a more nuanced picture of the effects of the intervention, we next study the heterogeneity in effects by prior academic performance and by gender.

4.3 Heterogeneity of effects

Prior evidence on online education suggests that its negative effects are larger for weak and for male students (e.g., Figlio et al., 2013; Xu and Jaggars, 2014; Bettinger et al., 2017). We therefore investigate the heterogeneity of our effects in Table 4. We start with the analysis by prior performance. Column (1) shows the impact on credits registered for. The interaction term is insignificant, but points towards a higher treatment effect for those with more credits in the winter term. The interaction in Column (2) shows that students with better prior performance benefit more from the program in terms

Table 3: Average Impacts of Online Peer Mentoring on Student Outcomes

Dependent Variable:	Credits				GPA	
	Registered		Earned			
	OLS	IV	OLS	IV	OLS	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, estimating intent-to-treat effects. The even-numbered columns instrument a dummy for initial program take-up by the (random) treatment assignment variable, estimating treatment-on-the-treated effects. 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. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

of credits earned. The point estimates suggest a positive effect starting at around 23 credits (about five of the six scheduled courses) passed in the first term. There are no effects on GPA (Column 3).[16] Overall, those who fared well in the winter term benefited from the mentoring program. In contrast, weak students do not seem to benefit from the program. This is interesting because in several evaluated (peer) mentoring programs in higher education, good students are excluded (e.g., Angrist et al., 2009).[17] These results also raise the question whether similar patterns (more able students benefiting more from the program) are also observed in the survey.

[16]Appendix Figure C.1 further illustrates this heterogeneity in an exploratory analysis. It shows the share of students who reach the 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. There is no change in the probability of having reached the goal in the lowest and middle tercile. For the highest tercile, there is a difference in the probability of reaching the study plan goal of almost 10 percentage points (p-value=0.059) or around 18% of the control group mean.

[17]We hired as mentors students with above-average performance in the same study program as their mentees. Hence, the heterogeneity in our data is in line with prior evidence suggesting that students perform better when being taught by similar persons (e.g. Dee, 2005; Hoffmann and Oreopoulos, 2009).

In Online Appendix B.4, we show heterogeneity analyses by credits earned in the winter term for our survey outcomes. While the results are more noisy, the overall pattern is similar. This bolsters our confidence that the heterogeneous effects by prior performance reflect a meaningful difference between more and less able students' response to the mentoring.[18]

We then turn to effects by gender. Column (4) shows a positive treatment effect for men, who register for around 2.7 more credits (> 0.5 additional courses) when offered treatment. The interaction is negative and of around the same magnitude, suggesting that female students do not benefit from the program. Column (5) shows similar results for credits earned. The results are again attenuated, with an effect of around one more credit earned by male students and zero effects for female students. We again find no effects on GPA. Again, both female and male students benefit more when they passed more credits in the winter term (not shown). This pattern is more pronounced for male students. These results are somewhat in contrast to prior evidence on mentoring programs. Angrist et al. (2009) find that an in-person program combining academic counseling with financial incentives positively affected female college students, with no effects on male students. In our context male students benefit more from the mentoring program. 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).

5 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 a more advanced online mentor. The structured one-on-one mentoring focused on study behavior, 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 German public university, where the entire summer term was conducted online.

We document three sets of main results. First, the peer mentoring program positively affected the students' motivation and study behavior. Second, while the impacts on study behavior and motivation translate into an increase of exam

[18] We show exploratory analyses in the Online Appendix. In Online Appendix C.4, we follow Abadie et al. (2018) and Ferwerda (2014) and estimate effects using endogenous stratification approaches. In line with the analysis above, students in the upper part of the distribution of predicted outcomes in the summer term seem to benefit most from the program.

Table 4: Treatment Effects by Student Characteristics

	By prior performance			By gender		
	Credits		GPA	Credits		GPA
	registered	earned		registered	earned	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.55	-4.13***	0.14	2.67***	0.91	0.03
	(2.92)	(1.57)	(0.21)	(0.99)	(0.83)	(0.07)
Treatment · credits WT	0.08	0.18***	-0.00			
	(0.10)	(0.06)	(0.01)			
Treatment · female				-2.73*	-0.79	0.01
				(1.40)	(1.22)	(0.10)
Mean dep.	26.33	17.66	2.52	26.33	17.66	2.52
Obs.	691	691	595	691	691	595

Note: This table shows impacts of peer mentoring on administrative student outcomes by prior performance and gender adapting equation 1. Columns (1) to (3) estimate interactions by prior performance, using students' credits earned in their first term as the meausre of prior performance. Columns (4) to (6) use interactions by gender. Columns (1) and (4) use the number of credits for which students registered in the summer term 2020 as the dependent variable. Columns (2) and (5) use the number of earned credits in the summer term as the dependent variable. Columns (3) and (6) use students' average GPA (running from 1=worst to 4=best) among earned credits in the summer term as the dependent variable. Standard errors are robust. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

registrations, the average treatment effect on passed credits is small and not significantly different from zero. Similarly, the students' GPA is not affected by our intervention. Third, across various outcomes, we observe a consistent pattern of heterogeneity in the treatment effects that stands in contrast with expectations on the impacts of peer mentoring. While students in the bottom part of the distribution of prior performance in the first term seem to be largely unaffected by the treatment, we observe a positive effect on students who performed well in their first term. Male students also benefit somewhat more from the program.

Our results provide the first evidence on the effectiveness of 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 could improve student outcomes even more.

References

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

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.

Angrist, N., P. Bergman, and M. Matsheng (2020): "School's Out: Experimental Evidence on Limiting Learning Loss Using 'Low-Tech' in a Pandemic," *NBER Working Paper No. 28205*.

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.

Barsbai, T., V. Licuanan, A. Steinmayr, E. Tiongson, and D. Yang (2020): "Information and the Acquisition of Social Network Connections," *NBER Working Paper No. 27346*.

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., L. Fox, S. Loeb, and E. S. Taylor (2017): "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.

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.

Clark, D., D. Gill, V. Prowse, and M. Rush (forthcoming): "Using Goals to Motivate College Students: Theory and Evidence from Field Experiments," *Review of Economics and Statistics*.

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.

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.

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*.

Hess, S. (2017): "Randomization inference with Stata: A guide and software," *Stata Journal*, 17, 630–651.

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.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007): "Experimental Analysis of Neighborhood Effects," *Econometrica*, 75, 83–119.

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*.

Kofoed, M., L. Gebhart, D. Gilmore, and R. Moschitto (2021): "Zooming to Class?: Experimental Evidence on College Students' Online Learning During Covid-19," *IZA Discussion Paper No. 14356*.

Lavecchia, A., H. Liu, and P. Oreopoulos (2016): "Behavioral Economics of Education: Progress and Possibilities," in *Handbook of the Economics of Education*, ed. by E. A. Hanushek, S. Machin, and L. Woessmann, Elsevier, vol. 5, 1–74.

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. W. Patterson, U. Petronijevic, and N. G. Pope (forthcoming): "Low-Touch Attempts to Improve Time Management among Traditional and Online College Students," *Journal of Human Resources*.

Oreopoulos, P. and U. Petronijevic (2018): "Student Coaching: How Far Can Technology Go?" *Journal of Human Resources*, 53, 299–329.

——— (2019): "The Remarkable Unresponsiveness of College Students to Nudging And What We Can Learn from It," *NBER Working Paper No. 26059*.

Orlov, G., D. McKee, J. Berry, A. Boyle, T. DiCiccio, T. Ransom, A. Rees-Jones, and J. Stoye (2020): "Learning During the COVID-19 Pandemic: It Is Not Who You Teach, but How You Teach," *NBER Working Paper No. 28022*.

Patterson, R. W. (2018): "Can Behavioral Tools Improve Online Student Outcomes? Experimental Evidence from a Massive Open Online Course," *Journal of Economic Behavior & Organization*, 53, 293–321.

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.

Scrivener, S., M. J. Weiss, A. Ratledge, T. Rudd, C. Sommo, and H. Fresques (2015): "Doubling Graduation Rates: Three-Year Effects of CUNY's Accelerated Study in Associate Programs (ASAP) for Developmental Education Students," *MDRC*.

Sommo, C., D. Cullinan, M. Manno, S. Blake, and E. Alonzo (2018): "Doubling Graduation Rates in a New State Two-Year Findings from the ASAP Ohio Demonstration," *MDRC Policy Brief 12/2018*.

Weiss, M. J., A. Ratledge, C. Sommo, and H. Gupta (2019): "Supporting Community College Students from Start to Degree Completion: Long-Term Evidence from a Randomized Trial of CUNY's ASAP," *American Economic Journal: Applied Economics*, 11, 253–297.

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.

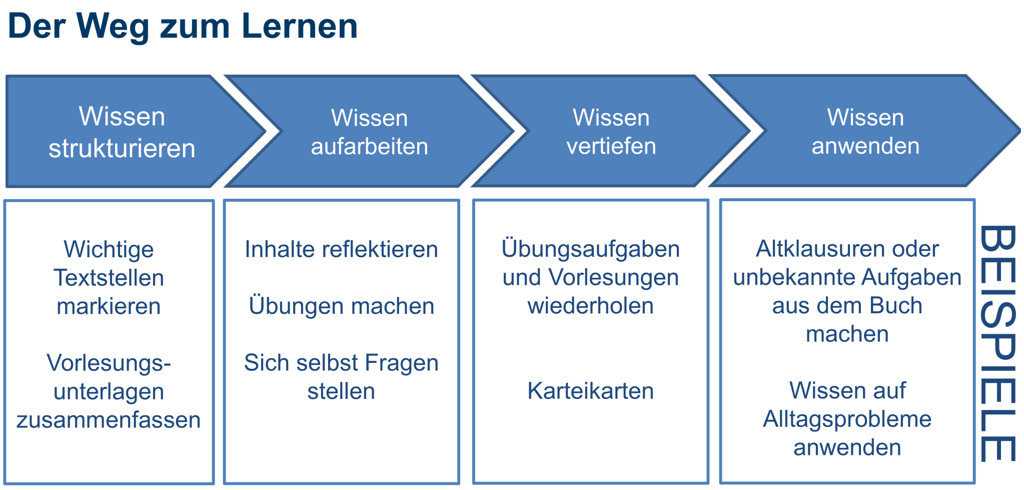
Figure A.1: Example: Input for Weekly Study Plan

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.2: Example: Actual 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.3: Example: Input on How to Study Effectively



A.2 Sorting into Mentoring Take-Up

Table A.1: Summary Statistics by Take-Up

	No take-up	Take-up	Difference
Female	0.43	0.54	0.10*
	(0.50)	(0.50)	(0.05)
Age	20.95	21.69	0.73**
	(2.45)	(2.94)	(0.29)
High-school GPA	2.37	2.39	0.02
	(0.58)	(0.66)	(0.07)
Top-tier high-school type	0.81	0.65	-0.16***
	(0.39)	(0.48)	(0.05)
Foreign univ. entrance exam	0.06	0.11	0.05*
	(0.24)	(0.32)	(0.03)
Earned credits in first term	25.29	25.20	-0.09
	(9.42)	(8.21)	(0.98)
First enrollment	0.69	0.67	-0.02
	(0.46)	(0.47)	(0.05)
Part-time student	0.10	0.06	-0.05
	(0.31)	(0.23)	(0.03)
Obs.	202	142	344

Note: This table shows means of administrative student data (standard deviations in parentheses) by take-up among treated students, together with differences between means and corresponding standard errors (in parentheses). In the line where we report high-school GPA we need to drop 6 observations where we do not have information on students' high-school GPA.

A.3 Mentoring Differences by Mentor Gender

In this subsection, we provide descriptive evidence on differences in mentoring by mentor gender. We use only those students who signed up for the program, an endogenous outcome itself.

Table A.2 shows that 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.

Table A.3 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.2: Meetings by Mentor Gender

	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 A.3: 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 Survey Information and Evidence

B.1 Survey Questions and Sorting into Survey Participation

In table B.1, 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, and how likely they think they will graduate in time. With the exception of these questions, all responses are measured on a five-point Likert scale where higher values indicate higher agreement with the question. Table B.2 shows that while survey participants slightly differ from non-participants on their observable characteristics, participation is balanced across treatment and control group.

Table B.1: Survey Questions

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 B.2: Sorting into Survey Participation

	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

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

B.2 Regression Results

Table B.3: Treatment Effects on Assessment of Own Motivation and Study Effort

Dep. Var.:	Motivation		Continuous studying		Timely exam prep.		Sufficient effort	
	OLS	IV	OLS	IV	OLS	IV	OLS	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, estimating intent-to-treat effects. The even-numbered columns use (random) treatment assignment variable as an instrument for initial program take-up, estimating treatment-on-the-treated effects. 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 B.4: 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	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	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, estimating intent-to-treat effects. The even-numbered columns use (random) treatment assignment variable as an instrument for initial program take-up, estimating treatment-on-the-treated effects. 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 B.5: 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	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	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, estimating intent-to-treat effects. The even-numbered columns use (random) treatment assignment variable as an instrument for initial program take-up, estimating treatment-on-the-treated effects. 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. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

B.3 Average Standardized Effects

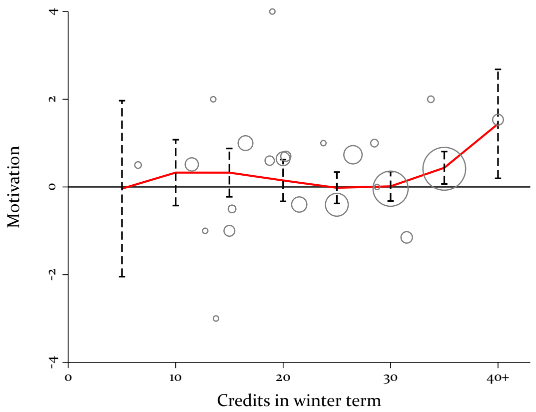
Table B.6: Average Standardized Effects on Survey Outcomes

	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

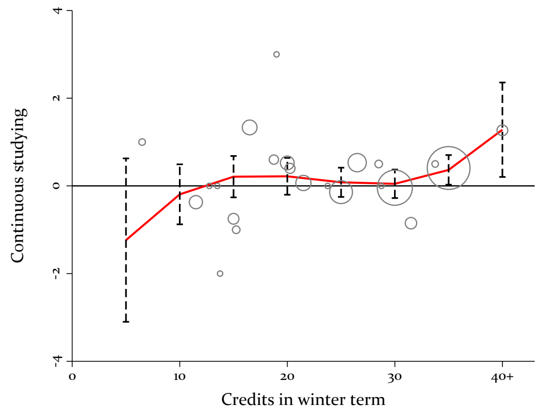
Note: This table shows average standardized effects on survey outcomes by broad group using the methodology of Kling et al. (2004) and Clingingsmith et al. (2009). Standard errors are robust. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

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

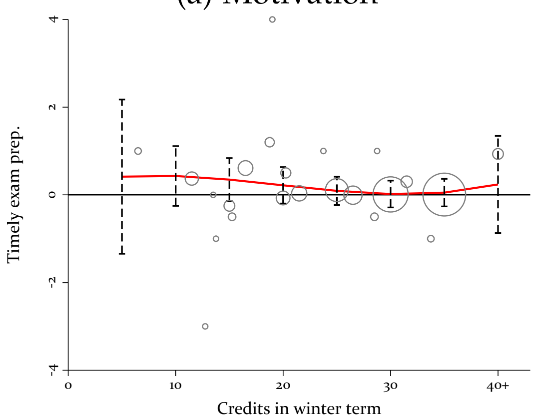
Figure B.1: Treatment Effects on Survey Outcomes by Credits Earned in Winter Term



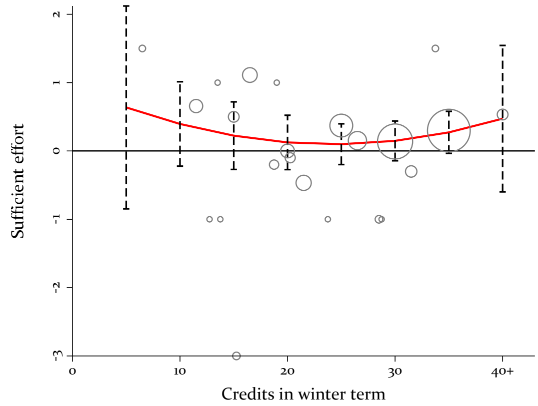
(a) Motivation



(b) Continuous studying



(c) Timely exam prep.



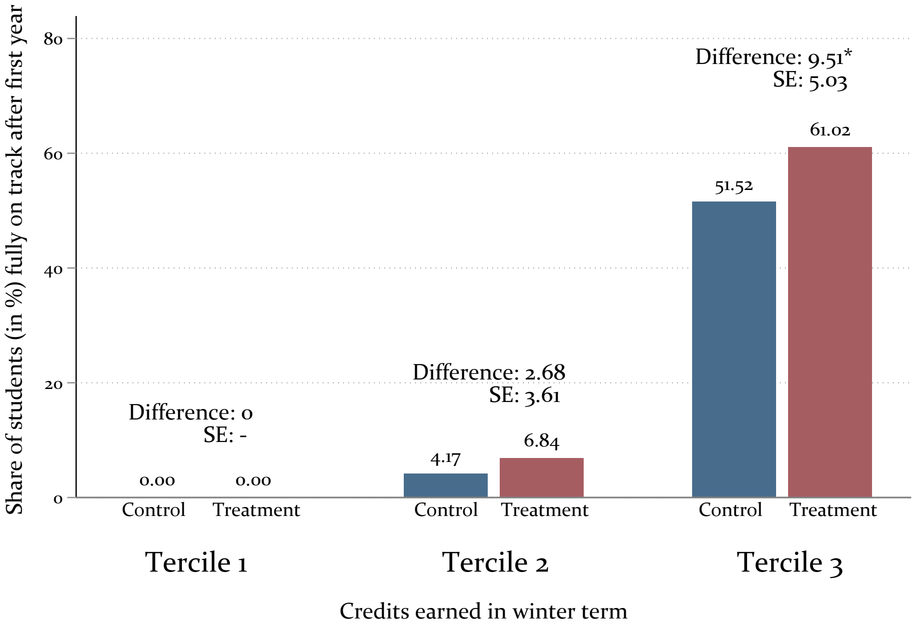
(d) Sufficient effort

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.

C Additional Results for Administrative Student Outcomes

C.1 Effects on Reaching First Year Goal

Figure C.1: 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. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

C.2 Heterogeneity by High School GPA

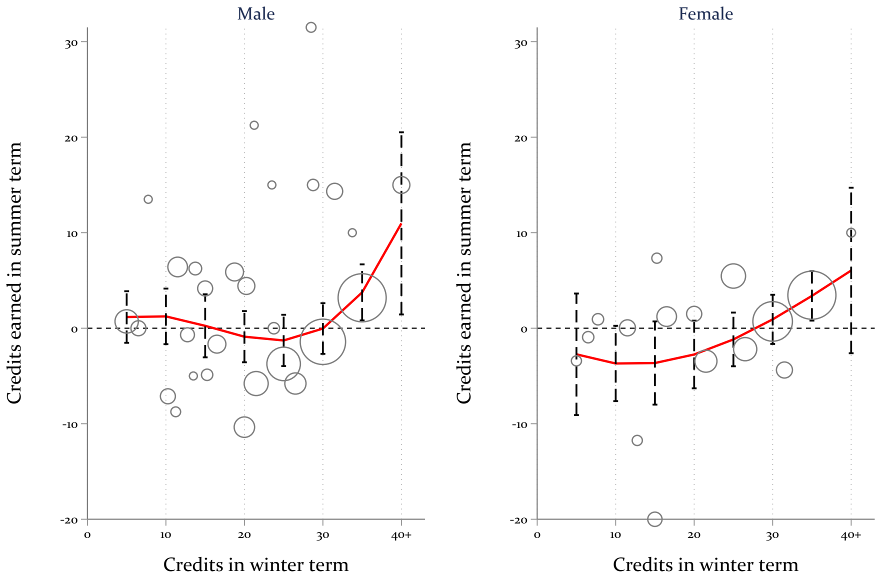
Table C.1: Treatment Effects by Students' High School GPA

	Credits		GPA
	registered	earned	
	(1)	(2)	(3)
treatment==1	2.82	-9.94***	-0.67***
	(2.33)	(1.71)	(0.16)
Treatment · H.S. GPA	-0.64	4.37***	0.29***
	(0.88)	(0.67)	(0.06)
Mean dep.	26.37	17.68	2.52
Obs.	680	680	586

Note: This table shows impacts of peer mentoring on administrative student outcomes by high school GPA adapting equation 1. Column (1) uses the number of credits for which students registered in the summer term 2020 as the dependent variable. Column (2) uses the number of earned credits in the summer term as the dependent variable. Column (3) uses students' average GPA (running from 1=worst to 4=best) among earned credits in the summer term as the dependent variable. Standard errors are robust. * p<0.10, ** p<0.05, *** p<0.01

C.3 More Heterogeneity by Gender

Figure C.2: 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.

C.4 Endogenous Stratification

Table C.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.

C.5 Mentor "Value-Added"

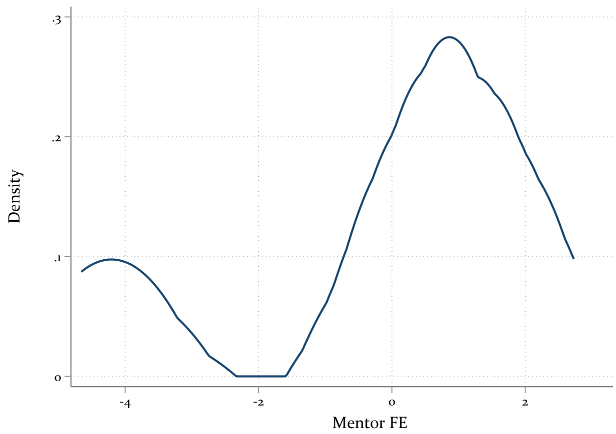
In this subsection, we provide results from a sort of "value-added" analysis where we regress students' outcomes in the summer term on their performance in the winter term, their observable characteristics, and a mentor dummy. We label the estimated mentor fixed effect in each outcome dimension the mentor's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

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

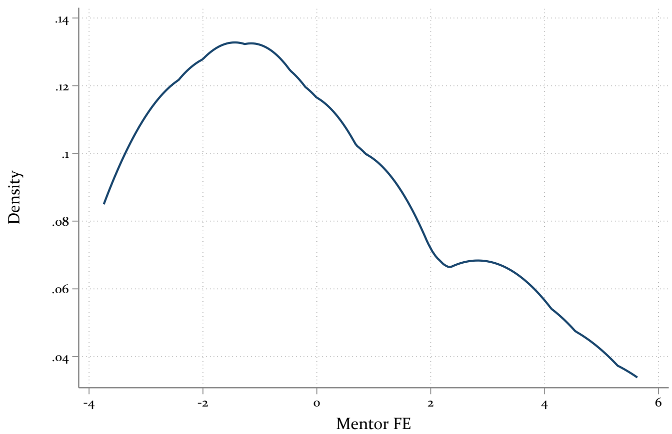
where y_i is the respective outcome of student i in the summer term, X_i is a vector of observable characteristics including past performance (credits earned in the winter term and high-school GPA), students' gender and age, a dummy whether students are from the region where the university is located, 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 C.3 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 C.3. 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.

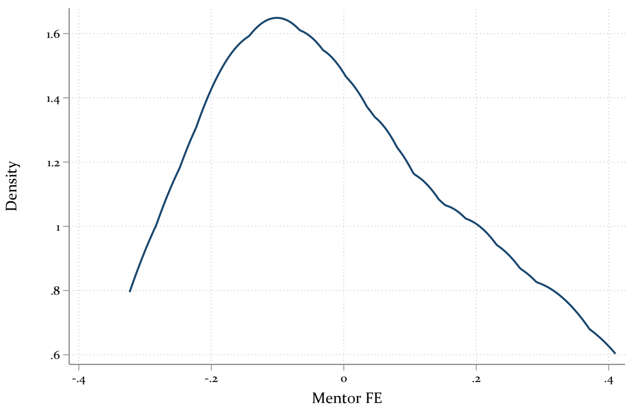
Figure C.3: Kernel Density Plots of Mentor Value-Added



(a) Credits registered for



(b) Credits earned



(c) GPA

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.

Table C.3: Pairwise Correlations of Mentor Value-Added

	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

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