Can peer mentoring improve online teaching effectiveness? An RCT during the COVID-19 pandemic

David Hardt\(^a\), Markus Nagler\(^b\), Johannes Rincke\(^c\)

\(^a\) University of Erlangen-Nuremberg, Germany
\(^b\) University of Erlangen-Nuremberg, CESifo, and LASER, Germany
\(^c\) University of Erlangen-Nuremberg and CESifo, Germany

**A R T I C L E I N F O**

**Keywords:**
Peer Mentoring
Online Education
COVID-19
Teaching Effectiveness
Higher education

**A B S T R A C T**

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 there is evidence for positive effects on the most able students.

1. Introduction

Online delivery of tertiary education has taken center stage. The COVID-19 pandemic has forced virtually all education institutions to switch to online teaching. However, the literature has generally found online teaching to be inferior to classroom-based teaching (e.g., Bettinger et al., 2017; Figlio et al., 2013). 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., 2022; 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 spring term 2020. Our sample comprises 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 that we collected before the start of the examination period in the spring term 2020. 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 topics from prior meetings. 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. Including one additional mentee into the program for a three-month period would cost about € 60.

\footnote{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. We thank the editor, Margherita Fort, as well as three helpful referees for comments and suggestions. Ushi Backes-Gellner, Christine Binzel, Ben Castelman, Fabian Dehos, Ulrich Glogowsky, Joshua Goodman, Philipp Lergetporer, Claus Schnabel, Jeffrey Smith, Martin Watzinger, 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 Andersen, Helen Dettmann, Eva-Maria Drasch, Daniel Geiger, Nina Graßl, Jana Hoffmann, Lukas Klostermeier, Alexander Lempp, Lennart Mayer, Jennifer Meyer, Annabel Pepp, 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.}

\footnote{Corresponding author.}

\footnote{E-mail addresses: david.hardt@faue.de (D. Hardt), markus.nagler@faue.de (M. Nagler), johannes.rincke@faue.de (J. Rincke).}

https://doi.org/10.1016/j.labeco.2022.102220

Received 26 October 2021; Received in revised form 27 June 2022; Accepted 30 June 2022

Available online xxx

0927-5371/© 2022 Elsevier B.V. All rights reserved.

Please cite this article as: D. Hardt, M. Nagler and J. Rincke, Can peer mentoring improve online teaching effectiveness? An RCT during the COVID-19 pandemic, Labour Economics, https://doi.org/10.1016/j.labeco.2022.102220
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 terms, students are to pass six courses each worth five credits. Since the second term includes more rigorous courses relative to the first term, many students struggle in this term.1 A key advantage of our setting is that the spring term 2020 was conducted entirely online because the German academic spring term starts and ends later (April to July) than is common internationally. Thus, there are no spillovers from in-person to online classes.

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 spring term 2020, and are more likely to state they provided enough effort to reach their goals in this term. In contrast, students’ views on departmental services or on online teaching in the spring term 2020 or in general seem unaffected. 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 expectations based on prior research suggesting that weaker students struggle most in online learning (e.g., Bettinger et al., 2017; Figlio et al., 2013) 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 effectiveness of online education. This literature has found online teaching to be less effective than classroom-based teaching (Bettinger et al., 2017; Figlio et al., 2013), likely due to problems of disorganization among students in online teaching (e.g., Banerjee and Duflo, 2014). Research on specific aspects of the online education production function is scarce. Closest to our work, Oreopoulos and Petronijevic (2022) 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 role of peer mentoring programs for online higher education.

We also contribute to the experimental literature on mentoring interventions in higher education.2 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 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., 2022; Bacher-Hicks et al., 2020; Grewenig et al., 2021). The closest paper is Carlana and La Ferrara (2020), who experimentally assigned struggling Italian middle school students an online tutor during the pandemic and report positive effects on performance and well-being. They specifically target students from less affluent backgrounds. Our paper contributes by studying the effectiveness of online mentoring in higher education. Our strongest effects are for already well-performing students, indicating that peer mentoring in online education may increase inequality in outcomes. This is very different from tutoring, where we find a reduction in inequality in a follow-on project in a similar setting (Hardt et al., 2022). Despite the nearly 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 in spite of emerging evidence showing that the shift led to worse learning outcomes and depressed student expectations and outlooks (Altinding et al., 2021; Bird et al., 2022; De Paola et al., 2022; Jaeger et al., 2021; Kofod et al., 2021; Rodriguez-Planas, 2022a; 2022b).

We acknowledge that the situation during the pandemic was special relative to regular online education settings. We nevertheless believe that our results carry lessons beyond the pandemic and discuss this in more detail in Section 2.

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 term. Administrative data show that large shares of students do not complete 30 credits per term, 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.3 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 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.

Most likely, given prior findings on such problems in online education, these issues were exacerbated by the switch to online teaching. In addition, online education also may have affected other margins, such
as instruction quality, peer interactions, differences in the study environment, distractions, and motivation. Given our program content, the mentoring intervened with students’ self-organization, helped them to study more efficiently, and affected students’ handling of distractions as well as their motivation.

Due to the COVID-19 pandemic, in the spring 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 spring term 2020 was conducted entirely online because the German academic year starts and ends later than common internationally. It is therefore cleaner than would be possible in similar settings during the pandemic since there are no spillovers from in-person to online teaching.

We leveraged the COVID-19 pandemic to assess the effectiveness of mentoring programs when students lack formal in-person interaction at university. However, in many respects, the pandemic situation was different than other online education settings. First, instructors did largely not have much experience teaching virtually via Zoom (Altindag et al., 2021; Orlov et al., 2021). Second, students experienced a variety of negative shocks, from cancelled internships to adverse labor market and health events for their family members (Jaeger et al., 2021). Third, Germany went into a partial lockdown from end of March to early May 2020. Thus, students may have suffered from isolation and depression more than in other online education settings. Finally, the mentors themselves may have also been affected by the challenges that came with the pandemic. We still believe that our results carry implications beyond the pandemic. First, we do not observe strong impacts on some survey questions that should capture dimensions of such pandemic effects. For instance, treated students do not report being in touch with peers more than control group students or do not feel differentially valued by the department. Second, similar impacts on student motivation and study behavior are found in the literature analyzing comparable interventions outside pandemics (Oreopoulos et al., 2020). Third, if anything, the fact that we implemented our intervention during a partial lockdown should make the treatment more effective than typical interventions, particularly for those students most affected by the pandemic (Rodriguez-Planas, 2022a; 2022b). However, we find similarly muted effects on average than mentoring interventions outside the pandemic (e.g., Oreopoulos and Petrojenic, 2019) and show that the intervention had the strongest impact on students who already performed well before the onset of the pandemic. We also do not see differential impacts on students who had greater local support as proxied by being from the region around the university.

2.2. The mentoring program

In the first term week of the spring term 2020, students in the treatment group were informed via e-mail about the launch of a mentoring program for students in the second term. They were invited to register for the program through a webpage.7 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, well within our prior target length of 30–45 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 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 Figs. A.1 and A.2 in the Online 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., 2020). The third meeting focused on exam preparation (timing of exams, implications for mentees’ 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 Fig. A.3 in the Online 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., 2022). 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 summary, our mentoring program targets similar elements as the one studied by Oreopoulos et al. (2022). They find no effects of mentoring on student outcomes and only slight effects on study behavior. Relative to their intervention, our program is more intensive, offering regular contact and guidance, which may be required to shift academic outcomes. In line with this, Oreopoulos and Petrojenic (2018) find that among several student coaching methods, one-on-one peer mentoring was effective to raise student outcomes while technology-based interventions were ineffective. These elements may be particularly important in an online context.

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.9 All mentors were students who successfully completed the first year and were enrolled in the fourth term of the study program during the spring term 2020. They had good program 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 mentees were female and seven were male.10

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

---

7 Teaching started on April 20th and the exam period started on July 20th, 2020.
8 For more information, see https://www.wiso-virtuell.fau.eu/, last accessed October 19, 2021.
9 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.
10 Section A.3 in the Online Appendix shows that female and male mentors differ slightly in their mentoring. They do not seem to be differentially effective, though.
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, 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. 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 Fig. 1.

2.4. Sampling and random assignment to treatment and control group

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 term, who were not formally in their second term in the spring 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 winter term 2019/20, their first term at university.11 This leaves us with a sample of 694 students. Using a stratified procedure (strata variables: gender and earned credits in the winter term 2019/20), we randomly assigned all students in the sample to treatment or control group such that both groups were of equal size. We contacted all students in the treatment group via e-mail and offered them the chance to participate in the mentoring program. Students in the control group were not contacted and could not sign up for the program. After the intervention ended, we had to drop another three students from the sample who got credit for second-term courses earned elsewhere.12 Our final sample thus consists of 691 students (344 in treatment and 347 in control).13

We opted for a design that offers mentoring services to a random subset of the student population and against the alternative of an over-subscribed lottery for two main reasons. First, mentoring services are typically offered to students directly, without any pre-screening of students who might be interested in such services. Our design is thus a natural one and provides the most informative estimates. Second, in the specific situation of the pandemic and the uncertainties associated with it, we were concerned that an oversubscribed lottery could negatively affect students who in the end would not get an offer.

142 of the 344 students in the treatment group signed up for the mentoring. We randomly assigned these 142 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 students in the treatment group who signed up for the mentoring program, about 54 percent were female. Given that eight of the mentors were female and seven were male, 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 elicited the students’ assessment of their study effort, their satisfaction with the department’s effort to support online learning during the spring term 2020, views on online teaching generally, and beliefs about their academic achievement. The full set of questions is shown in the Online Appendix. 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 and participation was balanced across treatment and control group.

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 course that they pass) for which students register to measure attempted examinations, interpreted as student effort. Students do not have the right to participate in exams if they did not register in advance. Since the registration window typically closes after only half of the term passed, students have to commit rather early whether they want to take an exam. Second, our main outcome of interest is credits earned in the spring term 2020. This measures most directly the students’ academic achievement during the intervention term. This variable includes all credits earned in the spring term 2020, i.e., credits earned from exams that were supposed to be taken in the spring term 2020, credits earned from retaken exams that were supposed to be taken in the winter term 2019/20, and credits earned
from other exams.\textsuperscript{14} Note, however, that this might be a slow-moving variable since study effort has cumulative gains over time.

Students can in principle repeat each exam twice.\textsuperscript{15} Additional costs of not passing specific exams in the first attempt is that students must have passed 50 credits in total by the end of the third term. If they do not reach this goal, they need to leave the program. Due to the Covid-19 pandemic, these restrictions were loosened a bit in that attempts to pass an exam did not count towards the total number of attempts for each exam. In addition, students were given an additional term to reach the goal of 50 credits. Thus, the opportunity costs of not passing courses decreased in this term. However, note that this is true both for the treatment and the control group and should thus not affect treated students differentially.

Note that in the second term, each exam is worth 5 credits since each compulsory module is graded only on the final exam. However, in principle, courses may also be graded through exams covering parts of the course, worth less than the 5 credits the course is worth. For example, courses may require students to give a presentation in addition to the final exam, with the credits split between both required assessments. Many courses in the first term use this option, which is why students can earn credits that are not multiples of 5. This is why we use credits registered for and credits earned as dependent variables instead of using the number of courses or exams that students register for and pass.

Third, we examine the impact on students’ GPA for passed courses, running from 1 (passed) to 4 (best grade).\textsuperscript{15} 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 North America. 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 and missing GPA.

The exams took place after the end of the teaching period between end of July and September 2020. In addition to the data on exams, the registry data also contain 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 0 as the fail grade to 4 as the best grade).

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 part-time students.\textsuperscript{17} As can be seen from Table 1, the treatment and control groups were well balanced across all characteristics.

\textsuperscript{14} When we run the same analyses using only credits earned from exams that were supposed to be taken in the spring term 2020, the results follow the same pattern, but the effect on exams registered is not significant anymore.

\textsuperscript{15} In each term, students only have the opportunity to sit an exam once. The next opportunity to take the exam is then in the subsequent term.

\textsuperscript{16} 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.

\textsuperscript{17} 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.

\textsuperscript{18} To characterize compliers relative to the overall population in our sample, we also use the ivdesc command by Marbach and Haagartner (2020) and report results in Table A.2 in the same appendix. We find similar results. Comparing the impacts of sign-up on outcomes using OLS vs. IV in Appendix Table A.3 however suggests that compliers are positively selected on unobservables.

\textsuperscript{19} In unreported analyses, we find that the intensity of take-up (i.e., the extent to which students assigned to the treatment group actually made use of mentoring services, measured by the number of mentoring meetings attended) correlates significantly with some student characteristics. In particular, across various measures of actual service use, take-up is positively associated with high-

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Summary statistics by treatment status.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
</tr>
<tr>
<td>Female</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>(0.50)</td>
</tr>
<tr>
<td>Age</td>
<td>21.29</td>
</tr>
<tr>
<td></td>
<td>(2.48)</td>
</tr>
<tr>
<td>High-school GPA</td>
<td>2.37</td>
</tr>
<tr>
<td></td>
<td>(0.57)</td>
</tr>
<tr>
<td>Top-tier high-school type</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>(0.43)</td>
</tr>
<tr>
<td>Foreign univ. entrance exam</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>(0.25)</td>
</tr>
<tr>
<td>Credits VT</td>
<td>25.23</td>
</tr>
<tr>
<td></td>
<td>(9.27)</td>
</tr>
<tr>
<td>First enrollment</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>(0.48)</td>
</tr>
<tr>
<td>Part-time student</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>(0.28)</td>
</tr>
<tr>
<td>Obs.</td>
<td>347</td>
</tr>
</tbody>
</table>

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. Credits\textsubscript{IV} is the number of ECTS credits earned by the student in the winter term 2019/20.

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 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 of survey completion is unrelated to treatment assignment. Within the sample of participants, treatment and control group are balanced across all characteristics.

Take-Up

Our main measure of take-up is whether students signed up for the program. Out of the 344 students assigned to the treatment group, 142 signed up for the program. Table A.1 in the Online Appendix shows that these students are slightly older, more likely female, and more likely to have a foreign university entrance exam than students from the treatment group who did not sign up. The differences are small, however.\textsuperscript{18} Students who registered for the program could drop out at any time with no penalty. Table 2 shows the impact of offering treatment on the likelihood of sign-up, i.e., the first stage of our IV regressions.

As mentioned before, students assigned to the control group could not sign up for the program. Table 2 shows that from the 41\% of students in the treatment group who signed up for the program (Column 1), some students did not attend any meeting. 37\% of students in the treatment group took at least one meeting (Column 2). Further students leaving the program during the spring term 2020 reduced the share of students in the treatment group who took all five meetings to 32\% (Column 3).\textsuperscript{19}
Table 2

Take-up and first stage estimates.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Sign-up</th>
<th>Sign-up &amp; attended</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Treatment</td>
<td>0.41***</td>
<td>0.37***</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Underid. Test</td>
<td>179.60</td>
<td>158.01</td>
</tr>
<tr>
<td>Weak IV Test</td>
<td>241.30</td>
<td>203.70</td>
</tr>
<tr>
<td>Obs.</td>
<td>691</td>
<td>691</td>
</tr>
</tbody>
</table>

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 2019/20, i.e., the first stage of our instrumental variable regressions. (Column 1) uses initial program sign-up as the dependent variable and measure of take-up. This is our main take-up measure. Column (2) uses a dummy for having met at least once with the mentor as the dependent variable. Column (3) uses an indicator of whether students met five times with their mentors as the dependent variable. The underidentification and weak identification tests are the heteroskedasticity-robust Kleibergen and Paap (2006) rk LM and Wald F statistics, respectively, as reported by the `IVReg2` Stata command (Baum et al., 2007). Standard errors are robust. * p<0.10, ** p<0.05, *** p<0.01.

3.3. Estimation

To evaluate the intent-to-treat (ITT) effects of the peer mentoring program, we estimate the equation

\[ y_i = \alpha + \beta \text{Treatment}_i + \gamma_1 \text{Female}_i + \gamma_2 \text{CreditsWT}_i + \epsilon_i, \]

(1)

where \( y_i \) is the outcome of student \( i \), \( \text{Treatment}_i \) is an indicator for (random) assignment to the mentoring treatment group, \( \text{Female}_i \) is a dummy for female students, and \( \text{CreditsWT}_i \) is the number of (ECTS) credits earned by the student in the winter term 2019/20, the first term in which the students were enrolled. Each of the outcomes is thus regressed on the indicator for random assignment to the treatment group and the strata variables. We report robust standard errors.

Since treatment take-up was imperfect (i.e., not all treatment group students actually received mentoring services), we additionally run IV regressions using the random treatment assignment as an instrument for actual program take-up. Our main variable for measuring take-up is program sign-up (i.e., our first stage can be seen in Column 1 of Table 2). The first stage is expectedly strong, with a Kleibergen and Paap (2006) F statistic of around 240. The IV estimations essentially scale up the ITT point estimates of the treatment effect by the inverse of the sign-up rate.

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., Bettinger et al., 2017; Figlio et al., 2013). We thus expected heterogeneous effects by credits earned in the winter term 2019/20. 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 sign-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.5 in the Online Appendix, but find no differences. Second, the pre-analysis plan also specified that we would test if the effect on students enrolled at university 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. Fig. 2 shows results from ITT and from IV estimations (instrumenting sign-up by treatment assignment). We show the treatment effects along with 95% confidence intervals. All dependent variables are responses on a five-point Likert scale, with higher values indicating higher agreement with the statement.

Panel (a) shows treatment effects on students’ assessment of their motivation and study behavior in the spring term 2020. 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 economically and statistically significant effects on students’ response to whether they managed to study continuously throughout the spring term 2020. 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 (\( p\text{-value} = 0.048 \)). We believe that it is unlikely that these results merely reflect impacts on students’ mental health since in a more intensive tutoring intervention in later terms, we find no impacts on any of several dimensions of student mental health (Hardt et al., 2022). Thus, the treatment seems to have worked through affecting study motivation and study behavior.

In line with our expectations, 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 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

21 All corresponding tables are in the Online Appendix. The statements are also shown in the Online Appendix. Note that we treat the ordinal scales as if they were cardinal scales. Using ordered probit or ordered logit models renders qualitatively identical findings.

22 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 0.19 SD (\( p\text{-value} = 0.05 \)) and the treatment effect on the treated (TOT) shows an effect of 0.40 SD (\( p\text{-value} = 0.046 \)). Second, we use randomization inference procedures (Heil, 2017). Our standard errors are unaffected. Third, we use the procedure by Barsbai et al. (2020) that corrects for multiple hypothesis testing. Here, the \( p\text{-values} \) are larger, at around 0.15 for the questions that are significant above. While these results are thus a bit noisy, they are overall robust.

23 We show the heterogeneity of these effects by gender in Online Appendix B.S.
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. Again, we did not expect to see treatment effects on the respective outcomes, as the mentoring program did not target students’ general perceptions of online teaching. The specific items include students’ satisfaction with the departments’ online teaching content and implementation in the spring term 2020. 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 statistically and economically insignificant. We additionally analyze the students’ response to the question whether they frequently interacted with other students. This null result provides suggestive evidence that the program may not induce substitution for interaction among students but further research is needed to document whether the finding is robust.

We did not have baseline data and therefore did not pre-register minimum detectable effects. Considering a Z-score index (Kling et al., 2007) of all questions in the block of questions on motivation, continuous studying, timely exam preparation, and sufficient effort, we find that the treatment shifts the index by 0.19 standard deviations (p-value=0.05). The ex-post power for this effect is 0.66.

Overall, our results suggest that the peer mentoring program improved students’ motivation and study behavior, hence working as intended. One could be worried that these effects are driven by experimenter demand effects. It is therefore reassuring that we do not see effects on survey items which are unrelated to peer mentoring. In particular, the questions on students’ assessment of department services in general should suffer by experimenter demand effects, but we find no differences here. We also believe that this is not an issue since the survey was sent from the e-mail address of the dean of studies, a departmental e-mail address frequently used to inform students on departmental issues and to reach out to students. The survey was sent to all students and not targeted to treated students only. Finally, it asked students for their opinion on how the department fared in the spring term 2020. We thus believe that experiment demand effects are not an issue. We now investigate whether the positive effects on motivation and study behavior translated into improved academic outcomes.

4.2. Average impacts on primary outcomes

Table 3 shows differences between students in treatment and control group for academic outcomes.

The odd-numbered columns show ITT estimates, the even-numbered columns show corresponding IV estimates where we use the indicator for assignment to the treatment group 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

24 We additionally elicited students’ expectations of the likelihood of completing their studies on time and their planned credits. The results are noisy and show no difference between treatment and control group (not shown).

25 Using the latest available data from the early winter term 2021/22, when the students in our sample are in their fifth term of the program, we find that the program also fades out over time. Treated students register for more credits and earn more credits on average in the subsequent winter term 2020/2021.
Table 3
Average impacts of online peer mentoring on student outcomes.

<table>
<thead>
<tr>
<th></th>
<th>Credits</th>
<th></th>
<th>GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Registered ITT</td>
<td>Earned ITT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Treatment</td>
<td>1.39***</td>
<td>3.37**</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>(0.70)</td>
<td>(1.69)</td>
<td>(0.61)</td>
</tr>
<tr>
<td>Credits WT</td>
<td>0.28***</td>
<td>0.28***</td>
<td>0.81***</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Female</td>
<td>1.91***</td>
<td>1.74**</td>
<td>1.28**</td>
</tr>
<tr>
<td></td>
<td>(0.71)</td>
<td>(0.70)</td>
<td>(0.62)</td>
</tr>
<tr>
<td>Mean dep.</td>
<td>26.33</td>
<td>26.33</td>
<td>17.66</td>
</tr>
<tr>
<td>Underid. Test</td>
<td>179.6</td>
<td>241.3</td>
<td>179.6</td>
</tr>
<tr>
<td>Weak IV Test</td>
<td>691</td>
<td>691</td>
<td>691</td>
</tr>
</tbody>
</table>

Note: This table shows impacts of peer mentoring on administrative student outcomes using Eq. 1. The odd-numbered columns use OLS regressions, estimating intent-to-treat effects (labeled “ITT”). The even-numbered columns instrument a dummy for initial program sign-up by (the random) treatment assignment variable, estimating treatment-on-the-treated effects (labeled “IV”). Columns (1) and (2) use the number of credits for which students registered in the spring term 2020 as the dependent variable. Columns (3) and (4) use the number of earned credits in the spring term 2020 as the dependent variable. Columns (5) and (6) use students’ average GPA (running from 1=worst to 4=best) among earned credits in the spring term 2020 as the dependent variable. Credits WT is the number of (ECTS) credits earned by the student in the winter term 2019/20. The number of observations differs from Columns (1)-(4) since we have several students who do not earn any credits and thus do not have a GPA. The underidentification and weak identification tests are the heteroskedasticity-robust Kleibergen and Paap (2006) rk LM and Wald F statistics, respectively, as reported by the ivreg2 Stata command (Baum et al., 2007). Standard errors are robust. * p<0.10, ** p<0.05, *** p<0.01.

Although the estimates are imprecise. In the spring term 2021, these differences decrease. We also find that program dropout was unaffected. All results are available on request.

26 In Appendix Table A.3, we show the impact of sign-up on student outcomes without accounting for endogenous sign-up. When comparing OLS and IV results of this relationship, the OLS point estimates are larger throughout and significantly different from zero, but not significantly different from the IV estimates using versions of Hausman (1978) tests. The direction of changes in the point estimates however suggests that adopters are positively selected on unobservables. These results thus caution against analyzing impacts of mentoring without acknowledging endogenous participation decisions.

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., Bettger et al., 2017; Figlio et al., 2013; Xu and Jaggers, 2014). 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 2019/20. The interaction in Column (2) shows that students with better prior performance benefit more from the program in terms 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 winter term 2019/20. There are no effects on GPA (Column 3). This analysis suggests that highly motivated and highly able students, who often leverage the relatively easy first term to already take exams that they would have to do later (leading to credits in the winter term 2019/20 above 30) drive the heterogeneity effect shown in Table 4. These students comprise around 30% of our sample, making this behavior far from unusual in our setting.

Fig. 3 further illustrates this heterogeneity in an exploratory analysis. It shows credits earned by students by treatment status and by students’ tercile in the distribution of credits earned in the winter term 2019/20. The control mean is calculated as the students’ mean in the control group. The treatment mean is calculated as the control mean plus the estimated ITT effect, following Bergman et al. (2020). The fig-

27 Using machine-learning based methods to improve precision using more co-

variates (Upper, 2021; Wu and Gagnon-Bartsch, 2017) also does not meaning-

fully improve precision.

28 Taken at face value, the estimates suggest that the treatment effect is positive (at 0.41 for the average student in the treatment group who earned 25 credits in the winter term 2019/20. For students at the 25th percentile, the effect is slightly negative (-0.53), while for top performers at the 90th percentile we see strong positive effects (2.17). In Online Appendix C.5, we show flexible estimates of the relationship between credits earned in the winter term 2019/20 and treatment effects. We find that while the treatment is positive for good students, there is no evidence that it is detrimental for low performers.
Table 4
Intent-to-treat effects by student characteristics.

<table>
<thead>
<tr>
<th>By prior performance</th>
<th>By gender</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Credits</td>
</tr>
<tr>
<td></td>
<td>registered</td>
</tr>
<tr>
<td>Treatment</td>
<td>-0.55</td>
</tr>
<tr>
<td>(2.92)</td>
<td>(1.57)</td>
</tr>
<tr>
<td>Treatment × Credits WT</td>
<td>0.08</td>
</tr>
<tr>
<td>(0.10)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Treatment × Female</td>
<td>0.24***</td>
</tr>
<tr>
<td>(0.07)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Female</td>
<td>1.91***</td>
</tr>
<tr>
<td>(0.71)</td>
<td>(0.62)</td>
</tr>
<tr>
<td>Mean dep.</td>
<td>26.33</td>
</tr>
<tr>
<td>Obs.</td>
<td>691</td>
</tr>
</tbody>
</table>

Note: This table shows ITT estimations of the impact of peer mentoring on administrative student outcomes by prior performance and gender adapting Eq. 1. Columns (1) to (3) estimate interactions by prior performance, using students’ credits earned in their first term, the winter term 2019/20, as the measure 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 spring term 2020 as the dependent variable. Columns (2) and (5) use the number of earned credits in the spring term 2020 as the dependent variable. Columns (3) and (6) use students’ average GPA (running from 1=worst to 4=best) among earned credits in the spring term 2020 as the dependent variable. Standard errors are robust. * p<0.10, ** p<0.05, *** p<0.01.

Fig. 3. Treatment Effects by Tercile of Credits Earned in Winter Term 2019/20.
Note: This figure shows the number of credits students earned in the spring term 2020 by prior performance as measured by their tercile in the distribution of credits earned in the winter term 2019/20. 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 winter term 2019/20. The treatment mean is calculated as the control mean plus the estimated treatment effect. Standard errors reported are robust.

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., Angrist et al., 2009).29 These results also raise the question whether similar patterns (more able students benefiting more from the program) are also observed in the survey. In Online Appendix B.4, we show heterogeneity analyses by credits earned in the winter term

29
2019/20 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.30

We then turn to effects by gender. Column (4) of Table 4 shows a positive treatment effect for men, who register for around 2.7 more credits (> 0.5 additional courses) relative to students assigned to the control group. 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 2019/20 (not shown). This pattern is more pronounced for male students.31 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 spring term 2020 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 in exam 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: while students in the bottom part of the distribution of prior performance in the first term, the winter term 2019/20, seem to be largely unaffected by the treatment, we observe a positive effect on students who performed well in the winter term 2019/20. 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. Although we acknowledge that the pandemic situation may affect the effectiveness of such mentoring interventions, we believe that our results carry lessons beyond our setting. First, our intervention did not impact survey questions that were directly related to specific aspects of the pandemic, like being in touch with peers or feeling valued by the department. Second, our findings on student motivation and study behavior are well in line with previous literature analyzing comparable interventions outside pandemics. Third, a concern regarding external validity could be that our intervention would be more effective during a partial lockdown relative to normal times, particularly for those students most negatively affected by the pandemic. However, we find similarly muted effects on academic performance than mentoring interventions outside the pandemic on average, and specifically for students performing poorly at baseline.

Yet, we do believe that our intervention informs on some interesting aspects related to the pandemic and a potential recovery. Most importantly, students seem more motivated when interacting with a peer mentor, even though the interactions take place remotely and at low intensity. This is useful since remote interactions are viable even when the local supply of potential mentors is limited (Carlana and La Ferrara, 2020; Kraft and Falken, 2021). Our results also suggest that the program helps students to focus on their studies, which is one issue that many students struggled with during the pandemic. In the recovery from this pandemic, remote mentoring programs could thus help students in self-organizing better, being more motivated, and improving the transition from online to in-person teaching. We would like to highlight that, 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 more.

Regarding more general lessons on how to optimally design mentoring programs in higher education in the future, our results suggest that a “one-size-fits-all” approach may not be optimal. Most importantly, we document in our paper that the a low-dose mentoring intervention can significantly shift upwards the academic performance of the most able students, while the lower part of the ability distribution is unlikely to benefit. This pattern suggests that the type of mentoring we offered is a complement (rather than a substitute) to traits and skills that positively affect the students’ academic achievement in the first place (like motivation, conscientiousness, etc.). Starting off from recent work on focused vs. general interventions in K-12 education (Christensen et al., 2020), an interesting possible route for further research in higher education would be to test the effectiveness of targeted mentoring and/or tutoring interventions that allow to better address the individual needs of students. An optimal targeting of student support services may result in programs that offer mentoring services to well-performing students and more practical support services like tutoring to academically weaker students.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.labeco.2022.102220.

References


30 We show exploratory analyses in the Online Appendix. In the Online Appendix, we follow Abadie et al. (2008) 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 spring term 2020 seem to benefit most from the program.

31 In the Online Appendix, Fig. C1 shows heterogeneous treatment effects by credits earned in the winter term 2019/20, by gender. Again, both female and male students benefit more when they passed more credits in the winter term 2019/20. However, this pattern is much more pronounced for male students. Using credits registered as dependent variable shows similar effects (available on request).


