Pre-Analysis Plan: Can Peer Mentoring Improve Online Teaching? An RCT During the Corona Pandemic

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

May 18, 2020

Contents

[*]david.hardt@fau.de
[†]markus.nagler@fau.de
[‡]johannes.rincke@fau.de

1 Introduction

1.1 Abstract

We study a cohort of students at the School of Business, Economics, and Society at a German university in their second semester which is entirely held online due to the COVID-pandemic. We design and implement a program that provides students with a student advisor from a more advanced semester. The mentors and mentees meet one-on-one online and work on a plan for the semester, discuss issues in working from home or studying generally, and the mentors provide suggestions on how to study effectively. To determine whether and how students respond to the intervention, we measure the mentees' performance in exams.

1.2 Motivation

Online delivery of tertiary education is on the rise throughout the world. Currently, the Corona pandemic has forced virtually all education institutions to switch to online teaching. The literature on online teaching has generally found this format of teaching to be somewhat inferior to classical classroom based teaching [Brown and Liedholm, 2002; Figlio et al., 2013; Joyce et al., 2015; Alpert et al., 2016; Bettinger et al., 2017]. This may be due to problems of disorganization among students in online teaching, as has been argued for massive open online courses [so-called MOOCs; see e.g. Banerjee and Duflo, 2014; McPherson and Bacow, 2015]. However, students' lack of study skills may also be a contributing factor. Switching to online teaching therefore may aggravate a situation in tertiary education in which many students struggle to successfully complete their studies in time.[1]

One way how to improve on outcomes of online education could be to assist students in their self-organization by providing peer-to-peer mentoring. Leading universities like the MIT have launched student coaching programs to support their students during the COVID-19 crisis, and in May 2020 the American

[1]A large share of students never obtain a degree, and those who do often take much longer than the design of the program would suggest. For instance, data from the National Center for Education Statistics show that in the United States, less than 40 percent of a cohort entering four-year institutions obtain a bachelor's degree within four years. Data on other countries document that similar problems are widespread. Overall, in OECD countries the completion rate at the tertiary level is only 70 percent. See http://nces.ed.gov/programs/digest/d13/tables/dt13_326.10.asp. and https://www.oecd-ilibrary.org/education/education-at-a-glance-2013/indicator-a4-how-many-students-complete-tertiary-education_eag-2013-8-en

Economic Association (AEA) recommend that graduate programs should set up "more rigorous mentoring systems for students who will not be able to benefit from the usual sorts of interactions with peers and professors".[2] The literature on mentoring has so far mostly focused on settings before the onset of tertiary education [e.g., Lavy and Schlosser, 2005; Rodriguez-Planas, 2012; Oreopoulos et al., 2017]. In tertiary (classroom based) education, the results of mentoring interventions seem promising, although the literature is not large. Angrist et al. [2009] show that a combination of academic support services and financial incentives for good grades raises performance among female students. Bettinger and Baker [2014] show that a student coaching service focusing on aligning long-term goals and self-organization and providing study skills increased university retention. Recent evidence also shows that task-based goals are effective (and more productive than performance-based goals) in inducing student performance [Clark et al., forthcoming].

This trial is designed to test mentoring as a possible improvement in the effectiveness of online education for students in higher education. The context of the trial is the School of Business, Economics, and Society at a German university during the summer term 2020 that is taking place online due to the COVID-19 pandemic. In each fall semester, about 850 students enroll in the three-year (six-semester) bachelor's program *Economics and Business Studies*. This program is broad and can lead to specializations in business administration, economics, information systems, and business and economics education. The program requires students to collect 180 credit points to graduate. The study plan therefore assigns courses worth 30 credit points to each of the six semesters. In each of the first two semesters, students are supposed to pass exams in six compulsory courses, each of them worth five credit points. The specialization only starts after the first year in which students take compulsory modules.

Administrative data from the academic year 2018/19 shows that even in regular times, many students underperform relative to the suggested curriculum in the first study year: After the first semester, only 59 percent of students still enrolled at this point in time have completed courses worth at least 30 credit points. The curriculum for the second semester comprises some courses involving more rigorous methods relative to the first semester. As a result, students

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

typically further decrease their performance in the second semester: only about 25 percent have completed 60 credit points at the end of the second semester.

This study is meant to provide causal evidence on the effectiveness of a mentoring program in the second semester, especially given the fact that the online-teaching of courses due to the COVID-19 pandemic might exacerbate existing problems of self-organization and goal-setting. For this purpose, we designed a mentoring program that assigns students in their second semester a peer advisor currently enrolled in the fourth semester at the same school and who performed well in the first year of their university studies. The following sections present further details on our experimental design and the planned analysis of the data.

1.3 Research Questions

- Does mentoring improve the students' academic achievement in a context where all teaching is done online?
- Does the effect of mentoring on achievement differ by prior performance?
- Does the effect of mentoring on achievement differ by mentee gender?

2 Experimental Design

2.1 Intervention

The study program *Economics and Business Studies* at the university where the trial is going to be implemented requires students to collect 180 credit points to graduate. Students are expected to graduate after three years (six semesters). The study plan assigns courses worth 30 credit points to each semester. Administrative data show that a large share of students do not complete 30 credit points per semester, delaying their graduation. At the same time, survey data collected from an earlier cohort of students suggests that most students do not work full-time even if one aggregates the hours studied and the hours worked to earn income.[3] The salient study plan and target of achieving 30 credit points per term, the fact that most students do register for exams worth these credit points, and the fact that students do not seem to work enough to pass these exams

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

suggests that many students have problems in self-organizing and/or studying efficiently. This is where our program is supposed to intervene.

Due to the COVID-19 pandemic, in the summer term 2020 all courses of the School of Business, Economics, and Society will be conducted in online format. To this end, the university has acquired licenses of *Zoom*, an online video conference tool used widely in academic settings during this pandemic to digitize classes and seminars and to provide distance education. While the exact implementation of online teaching will differ by subject and instructor, this should make the setting similar to the setting of other academic institutions around the globe during this pandemic.

The trial focuses on the second semester consisting of six compulsory courses. We recruited 15 advisors who are themselves students in the *Economics and Business Studies* program at the School of Business, Economics, and Society. We hired students as advisors who successfully completed the first year of studies and during the current semester are enrolled in the fourth semester of the program.

In the first week of the semester, students in the treatment group are informed via email about the launch of a new mentoring program designed specifically for students in the second semester of the study program. They are invited to register for the program through a webpage.[4]

The mentoring program focuses on self-organization and is supposed to make mentees aware of potential problems and pitfalls of studying online. We designed the mentoring program to involve five one-on-one online meetings between advisors and mentees. Each meeting is supposed to last between 30 and 45 minutes. For each of the meetings, we provide advisors with structured information on how to conduct the session.

The first meeting is meant to focus on mentees' expectations regarding their performance in the second term, and to contrast this expectation with average performance figures from previous student cohorts. The advisor is also supposed to provide practical advice on how to self-organize when working from home. In the second meeting, advisors and mentees formulate specific goals that the mentee aims to achieve in the term. This includes aims regarding study effort (time schedule for the study week) and courses to be taken. It also includes performance-based goals (number of exams to pass). The third meeting is sup-

[4] The page asks for the students' consent to use their personal information for research purposes in anonymized form and for their consent to pass along their name and email address to their advisors. We sent reminder emails to students in the treatment group who did not register for the program within two days.

posed to focus on exam preparation (discuss timing of scheduled exams, reflect on implications for the mentee's preparation). The main topic of the fourth meeting is how to study effectively. This includes the presentation of a simplified four-stage learning model and how to implement the proposed learning strategies in practice. In the fifth and final meeting, the advisor and the mentee mainly discuss the mentee's exam preparation, including a time schedule that provides the mentee with guidance on how to specifically prepare for exams. In all meetings, besides the main topics mentioned, the advisor and the mentee are supposed to discuss current issues that the mentee is facing.

The advisors are asked to take brief notes about the content of the discussions during each meeting. We provide advisors with some structure for the notes in advance. Mentors are also instructed to prepare thoroughly for every individual meeting by recapturing the short notes they gathered during the prior meeting. To limit the risk of spillovers, we ask all advisors to make sure that the information is only provided to mentees and not to other students.

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

After the end of the exam period (preliminarily scheduled for July and August 2020), we will collect individual data on exam performance. We may also collect additional performance data for a further research paper or research note at a later point in time to assess long-run benefits of the program.

We do not expect that the School of Business, Economics, and Society will switch from online to classroom teaching during the semester and therefore plan for a full teaching period with online courses being the only or at least dominant way of teaching. However, if the overall situation changes significantly during the experimental period, we may allow advisors and mentees to meet in person for the meetings.

2.2 Hiring and Training of Advisors

For administrative reasons, we had to initiate the hiring of the student advisors about 4 weeks before the start of the program. In total, we hired 15 advisors. Work contracts were specified such that each advisor could handle a maximum of 10 mentees. The mentoring program's maximum capacity is therefore 150 students. All advisors were students who successfully completed the first year of studies and during the current semester are enrolled in the fourth semester of the study program.

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

On the next day, all advisors took part in a training given by professional coaches. The training focused on communications skills and took about five hours (excluding breaks). Three weeks after the start of the program, the advisors took part in a short supervision meeting (about one hour) with the coaches. In addition, the members of the research team sent regular emails to the advisors (one email before each of the five waves of meetings) and answered questions in response to individual queries by the advisors.

2.3 Data Collection

2.3.1 Administrative Data

We collect administrative data from the university to measure all outcomes related to exam participation and academic achievement. In addition, the university has provided us with background information on individual students. The individual characteristics include information on enrollment, gender, age, type of high school completed, and information on high-school GPA.

2.3.2 Survey Data

After the end of the intervention (i.e., after the fifth round of mentee-advisor meetings is completed), we will invite all students in the experiment to an online survey. The survey will be conducted using an existing platform at the department that is frequently used to survey students. Students who complete the survey will receive a payoff of € 8.00. The survey will elicit the students' satisfaction with the department's effort to support online learning during the teaching term, satisfaction with the one's own study effort, and beliefs about one's own academic achievement. For details, see Subsection 3.2.2.

2.4 Sampling

2.4.1 Randomized Treatment Assignment

About 850 students did enroll for the study program Business Studies for the fall semester of 2019. We excluded from the experiment students who dropped out after the first semester, who are not formally in their second semester, for example because of having been enrolled at another university before and having already completed courses from the first or second semester of the study program without having taken these exams at the university, and who completed less than a full course (5 credit points) in the first term.[5] This leaves us with 694 students entering the second term. We randomly assigned half of the students to treatment and the other half to control. The randomization was done in office by a computer. We used a stratified randomization scheme with gender and number of credit points completed in the first semester (three bins) as strata variables.

2.4.2 Invitations

To make sure that we do not overutilize the program's capacity, we first invited students sampled into treatment who did complete up to 30 credit points in their first term (369 students). We then successively invited three further groups of students sampled into treatment according to the number of credits points earned in the first semester, until all 347 students sampled into treatment got an invitation email.

In total, 140 students from the treatment group registered for the mentoring program. We may admit students who want to join later. Students registered for the program can drop out at any time with no penalty. We will drop students from the sample who are credited for courses in the second semester and earned the credits in an earlier term (either at the same university, or elsewhere). Such credits often show up with some delay in the administrative data. It is therefore possible that despite dropping students with such credits in the first semester, we have sampled some students who have already earned credit points for the second term.

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

2.4.3 Assignment of Registered Students to Advisors

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

2.5 Minimum Detectable Effects

We discuss minimum detectable effects for our main outcomes (total number of credit points earned in the second semester, and credit points attempted). We discuss minimum detectable effects for a significance level of 0.05 and a statistical power of 0.8.

From the baseline data for earlier cohorts, we expect the mean of credit points earned in the control group to be about 18.5 (SD 10.4). The minimum detectable intent-to-treat (ITT) effect is 2.2 credit points, or 21 percent of a standard deviation. We note, however, that the summer term 2020 differs from the baseline in that all courses are taught online. We cannot know whether (and how) the change in teaching mode affects the distribution of credit points earned absent mentoring. The true minimum detectable effect size might therefore differ significantly from the value provided above.

Due to the switch towards online teaching triggered by the COVID-19 pandemic, the university adjusted the rules regarding exam registrations. Normally, students have to sign up for exams beforehand, and they can sign out until three days before the exam date. Not showing up for an exam one has signed up for results in a 'fail' being recorded for the respective exam. For the summer term 2020, this rule was relaxed: students still have to sign up, but no-shows do not result in failing the exam. This change in the institutional environment can alter the students' decisions to take part in exams. It is therefore difficult to provide a minimum detectable effect for credit points attempted. If the students' behavior regarding exam participation is not altered relative to the baseline, the minimum detectable effect for credit points attempted is 2.1.

3 Empirical Analysis

3.1 Balancing Checks

We will check balance between treatment and control by t-tests (mean-comparison tests) on individual characteristics and by standardized differences. The characteristics included in the balancing checks will 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, credit points earned in the first term, a dummy for students who are in their first year at university, and a dummy for full-time students.[6]

We will run the same balancing checks on the sample of survey respondents. We will also study the selectivity in survey participation by means of mean-comparison tests between survey participants and non-participants.

3.2 Treatment Effects

3.2.1 Primary Outcome

Our primary outcome is credit points earned in the second term. This variable measures most directly the students' academic achievement during the term in which the intervention takes place.

We will also consider withdrawals during the second term (ending in September) as a possible outcome. If the effect of the intervention on withdrawals is insignificant, we will report this finding in the paper. Following Angrist et al. [2009], we will not exclude students who withdrew from the sample. Students who withdrew before earning any credit points in the second term will be coded as having zero attempted and earned credit points.

3.2.2 Secondary Outcomes

Program Take-up and Service Use A first set of secondary outcomes is meant to capture the decision to take up the offer to take part in the mentoring program. We will use the following variables:

- indicator for program sign-up

[6] About 10% of students are enrolled as part-time students because their university education is integrated into a vocational training program.

- indicator for students using some mentoring service (participation in at least one virtual meeting with the advisor)
- number of mentoring meetings completed
- indicator for students taking part in all five scheduled meetings with their advisor

Credit Points Attempted A further secondary outcome is credit points attempted in the second term. This variable measures the students' effort during the term in which the intervention takes place.

GPA We will also consider the second-term GPA as a secondary outcome. We note that this outcome is, in principle, affected by the student's decisions how many credit points to attempt and which exams to take. If we find that the effects on credit points attempted and earned are both insignificant, the effect on the GPA can nevertheless reveal a possible effect of the intervention on academic achievement.

Medium and Long-term Outcomes In addition to the outcomes for the second semester, we will later on also collect data on dropping out from the study program and graduation. As the study program is a three-year program and many students do not graduate in time, we will collect this data about four years after the beginning of the intervention. We may also collect data for medium-term outcomes (like credit points earned after the second study year). However, we do not commit to report those outcomes in the main paper. Hence, we may publish a paper on the short-term outcomes, and a separate paper (or a note) on the medium- or long-term outcomes.

Survey Outcomes From the student survey to be conducted after the end of the intervention, we will construct several additional outcomes:

- Satisfaction with the department's efforts to support the students' online learning (derived from 10-point Likert scale)
- Belief how strongly the department is devoted to individual students' academic achievement (derived from 10-point Likert scale)

- General assessment of online learning at university (derived from 10-point Likert scale)
- General preference for future role of online learning at university (derived from 10-point Likert scale)
- Rating of own study effort during the teaching term (derived from 10-point Likert scale, ranging from 'much less than usually' to 'much more than usually')
- Assessment of own ability to adapt to an online learning environment (derived from 10-point Likert scale)
- Expected number of credits earned in the second term (derived from a menu offering all first- and second term courses)
- Belief regarding the likelihood that one will graduate in time (i.e., after six terms/three years)

3.3 Estimation

To evaluate the treatment effects, we will run linear regressions. Each of the outcomes will be regressed on the treatment indicator and the vector of strata variables. We will report robust standard errors. However, not all students in the treatment group will take up the offer to receive mentoring services. Thus, in addition to intent-to-treat estimation regressions, we will run instrumental variable regressions using the randomized treatment assignment as an instrument for actual take-up. The first variable describing program take-up will be program sign-up. However, not all students who signed up will make use of mentoring services. We will therefore also estimate model variants where we use treatment assignment to instrument for actual service use.

For several reasons, we consider it likely that the treatment will have heterogeneous effects. A first observation is that prior evidence on online education shows that its negative effects are more pronounced among weaker students [e.g., Figlio et al., 2013; Bettinger et al., 2017]. We thus expect treatment effects to be negatively correlated with initial performance. This can be first measured through mentees' performance in the first term. Second, in the baseline, there is a positive correlation between the high school GPA and the probability to meet the 30 credit-points target in any term. This suggests to also use the high school GPA as a dimension to study the treatment effect heterogeneity by prior performance.

A second observation is that the literature has commonly found male students to suffer more from online relative to classroom education [e.g., Figlio et al., 2013; Xu and Jaggars, 2014]. At the same time, take-up rates in mentoring programs seem to be higher for female students [e.g., Angrist et al., 2009]. Thus, while we expect the effects of mentoring on outcomes among randomly chosen students to be larger for male than for female students, the relative magnitude of effects of having been offered an advisor on outcomes, and the relative effect of mentoring on outcomes conditional on take-up, is ex-ante unclear. We plan to study the effects by gender to inform on these questions.

A third dimension we plan to study is whether the effects of mentoring are larger when being mentored by female than by male advisors. Prior literature has found that interactions between student and instructor gender can matter for teaching effectiveness [e.g., Dee, 2005, 2007; Hoffmann and Oreopoulos, 2009]. As described above, we have made sure to have an around equal number of female and male advisors. In addition, due to the stratification of the sample by gender, we also have male and female mentees in the same proportions as in the underlying student population. Given the limited number of advisors, this analysis may however run into power issues.

Finally, we will also study if the treatment response of students enrolled at university for the first time differs from students who have been enrolled before. The rationale is that changing the study subject might reflect problems of the respective students to plan ahead and follow through with plans. We consider it unlikely that our treatment will be able to affect the behavior of those students. Again, given the limited number of students who are not enrolled for the first time, we may run into problems detecting any heterogeneity.

We plan to study the treatment effect heterogeneity by running regressions including an interaction term between the variable capturing the dimension of heterogeneity and the treatment indicator, together with the variable capturing the dimension itself. The strata variables will be included as controls.

The first two dimensions of a possible treatment effect heterogeneity will be reported in the paper. The other dimensions will be reported in the paper only if we find some heterogeneity. If not, the results will be relegated to the Online Appendix.

3.4 Other Variables

In addition to the analysis detailed above, we will also use the data provided by the notes that advisors take during each meeting. These notes may help us identify additional heterogeneity that is important to understand the effects of mentoring on academic performance in this setting. For example, research on MOOCs has found that students with problems in self-organization fare worse in online courses [Banerjee and Duflo, 2014]. These analyses will clearly be exploratory and it is difficult to gauge ex-ante which sort of information we will be able to obtain through these notes. Given the already quite structured nature of the mentoring, we however refrain from eliciting more details about the actual content of the mentoring sessions.

References

ALPERT, W. T., K. A. COUCH, AND O. R. HARMON (2016): "A Randomized Assessment of Online Learning," *American Economic Review: Papers & Proceedings*, 106, 378–382.

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

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

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.

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

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*.

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.

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.

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

JOYCE, T., S. CROCKETT, D. A. JAEGER, O. ALTINDAG, AND S. D. O'CONNELL (2015): "Does Classroom Time Matter?" *Economics of Education Review*, 46, 64–77.

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

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

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

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

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