The Persistent Effects of Short-Term Peer Groups in Higher Education

Thiemann, Petra

2018

Document Version:
Other version

Link to publication

Citation for published version (APA):
Thiemann, P. (2018). *The Persistent Effects of Short-Term Peer Groups in Higher Education*. (Working Papers; No. 2018:32).

Total number of authors:
1

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Working Paper 2018:32

Department of Economics
School of Economics and Management

The Persistent Effects of Short-Term Peer Groups in Higher Education

Petra Thiemann

November 2018

The Persistent Effects of Short-Term Peer Groups in Higher Education

Petra Thiemann*

Lund University & IZA

November 14, 2018

This paper studies the persistent effects of short-term peer exposure in a college setting. I exploit the random assignment of undergraduates to peer groups during a mandatory orientation week and follow the students until graduation. High levels of peer ability in a group harm the students' test scores and lead to increases in the probability of early dropout; this result is driven by the adverse effect of high-ability peers on low-ability students. I find suggestive evidence for discouragement effects: Peer ability is negatively correlated with the students' confidence in their academic ability after the first week.

JEL codes: I21, I23, J24
Keywords: peer effects, higher education, natural experiment, ability, educational attainment, dropout, major choice

*Contact: Lund University, Department of Economics, Tycho Brahes väg 1, SE-223 63 Lund, Sweden, Phone: +46 (0)46 222 8657, E-mail: petra.thiemann@nek.lu.se.
I am grateful to Bryan S. Graham and Michael Lechner for their invaluable advice while I was working on this project. I thank Olga Balakina, Jan Bietenbeck, Scott Carrell, Christina Felfe, Hans Fricke, Guido Imbens, Lance Lochner, Elena Manresa, Teresa Molina, Hanna Mühlrad, Alexandros Rigos, Michel Serafinelli, Kelly Shue, Andreas Steinmayr, and Ulf Zölitz as well as seminar participants in various places for their helpful comments. I thank Stephan Egger, Sharon Pfister, Jonathan Schulz, Uwe Sunde, Darjusch Tafreschi, and Christian Thöni for their help in the data collection. I thank the UC Berkeley Department of Economics for hosting me while working on this project. I acknowledge funding from the Swiss National Science Foundation (Grant no. 146969) and from the USC Dornsife Institute for New Economic Thinking.

1 Introduction

Peer effects play an important role in the design of university education, because interactions among students, both inside and outside the classroom, impact the students' performance and attainment (see Epple and Romano, 2011 and Sacerdote, 2011, for reviews). Principals and administrators therefore have to decide not only how many social interactions to build into the curriculum but also how to group students together in different contexts. This is especially important at the onset of university education—a phase when the students form habits and make decisions that impact their future careers.

A long-standing debate in this realm centers around ability peer effects: Does the presence of high-ability peers improve the performance and attainment of other students? The answer to this question is unclear from a theoretical perspective. On the one hand, positive spillovers can emerge in an environment where students learn from each other and where high-ability peers share their knowledge with other students (e.g., Hoxby and Weingarth, 2005, Oosterbeek and van Ewijk, 2014). On the other hand, high-ability students might discourage other students from providing effort by setting seemingly unattainable performance expectations (e.g., Rogers and Feller, 2016). Both hypotheses have found support in the recent empirical literature. While most articles find positive peer spillovers in university contexts (e.g., Carrell et al., 2009, 2013, Oosterbeek and van Ewijk, 2014, Feld and Zölitz, 2016), a smaller number of studies finds negative effects for subgroups of students (e.g., Carrell et al., 2013, Feld and Zölitz, 2016). These findings do not contradict each other because peer effects depend on the context under study. However, the question of which conditions give rise to either positive or negative spillovers remains open.

This paper presents new evidence on ability spillover effects in a typical peer intervention at the onset of university education—a mandatory orientation week for undergraduate students. Orientation weeks are implemented by colleges and universities around the world. They can generate meaningful spillovers because social ties are particularly malleable during the first days at college (Back et al., 2008). Furthermore, first interactions can shape the students' perception of their new environment. Studying an orientation week allows me to investigate the impact of a short-term intervention in the formation of peer groups on a large range of college outcomes up until the time of graduation: dropout decisions, course grades, choice of major, time to graduation, grade point average (GPA) after the first year, and final GPA.

To isolate the causal effect of peer ability from other factors that influence the students' performance and attainment, I exploit the random assignment of students to small groups during an orientation week at a Swiss university. The intervention leads to extensive interactions among peers because the students spend a 60-hour program in groups of 15 students. The main task during this week is an incentivized case study competition among the groups. The program thus promotes within-group interactions and leaves only limited time for between-group interactions.

The setting lends itself to the study of peer effects on dropout behavior and choice of major at college, two outcomes that have received little attention in the literature to date, compared to peer effects on GPA.[1] Dropout and major choice decisions are highly relevant from a labor market perspective because degree completion and the chosen major determine earnings in the long run (e.g., Altonji et al., 2016). In the present setting, the students have to decide on their major at the end of the first year, and they do not have to specify their courses at any point before that. Therefore, the students' choice of major can be influenced by the peers that they meet during the first year. Furthermore, the first year is characterized by high dropout rates. One-third of the students either drop out or fail in the first year; in the setting under study, more than 20 percent of the students drop out or fail during the first semester. Such dropout decisions might be influenced by peer groups as well (see Rogers and Feller, 2016, for peer effects in course completion).

The analysis draws upon a comprehensive administrative data set, which contains student background characteristics and detailed information on academic achievement (course grades, enrollment, dropout, and graduation records) for 8,000 undergraduates (cohorts 2003-2012).

[1] Few studies investigate peer effects on dropout behavior and major choice. Oosterbeek and van Ewijk (2014) and Booij et al. (2017) investigate the impact of college peer groups on dropout from college. Lyle (2007) investigates the impact of peer groups at the US Military Academy on the probability to remain in the Army. Brenøe and Zölitz (2018) investigate the impact of high school peers on selection into STEM majors.

I add two data sets to determine the impact of the freshmen week on friendship formation: an online survey inquiring about the persistence of students' friendships and study partnerships from the freshmen week, and data on student participation in mandatory "tutorial groups." Tutorial groups are small, self-selected groups of around 15 students, in which the students review the material of a course together with an instructor. The selection into these groups is informative about the persistence of first-week contacts.[2] Moreover, to investigate discouragement effects as a potential channel of peer effects, I use data on the students' confidence in their own performance from a survey and an experiment, which were carried out shortly after the intervention finished.

The results are as follows. In the first, descriptive, part of the paper, I document that the randomized orientation week groups have a persistent impact on the formation of friendships and study partnerships. Up to three years after starting their undergraduate degree, teammates are significantly overrepresented among a student's friends, according to self-reported friendship data. Moreover, teammates are significantly more likely to select into tutorial groups with other teammates, compared to students who were not teammates.

In the second part of the paper, I explore the causal effect of peer ability on student achievement. I find that higher levels of peer ability lead to lower grades and to higher dropout probabilities during the first semester. Both of these effects are driven entirely by the adverse impact of peer ability on low-ability students. Moreover, higher peer ability leads to a lower probability of choosing a double major. As a longer-term effect, I find a negative peer ability effect on final GPA. Using the data on the students' confidence that was collected shortly after the orientation week, I find a negative relationship between students' confidence levels and the level of peer ability in the group.

This paper makes two main contributions to the literature. First, the paper demonstrates negative causal effects of peer ability on a range of performance and retention outcomes. It thus contributes to the emerging evidence on negative ability peer effects at different levels of schooling (c.f. Carrell et al., 2013, Antecol et al., 2016, Feld and Zölitz, 2017).[3] The

[2]Enrollment into the tutorial groups is mandatory, but the students can choose their group freely. The groups differ by time of day and instructor.

[3]To frame the results in a positive way, students can benefit from being among low-ability peers—at least

findings are also in line with causal evidence on discouragement effects from the psychology literature (e.g., Rogers and Feller, 2016). In the present study, such discouragement effects appear most important for low-ability students and during the initial stages of university education, when the students slowly get to know their new environment and decide whether to stay or to drop out.

Second, the paper shows that short-term peer interventions matter. Prior studies have focused on longer-term interventions, which typically last for one semester or year. These interventions include assignments to work groups, study groups, dorm rooms, and military squadrons (Sacerdote, 2001, Stinebrickner and Stinebrickner, 2006, Lyle, 2007, 2009, Carrell et al., 2009, 2013, Oosterbeek and van Ewijk, 2014, Booij et al., 2017, Feld and Zölitz, 2017). The paper demonstrates that a short intervention can have long-lasting effects; therefore, administrators and universities should take such short-term interventions into account when considering how to improve student outcomes.

The remainder of the paper is structured as follows. Section 2 describes the empirical setup and data. Section 3 presents the dataset, and Section 4 details the empirical analysis. The results are presented in Section 5, followed by tests for their robustness in Section 6. Section 7 summarizes and discusses the findings.

2 Setting

2.1 The orientation week

The University of St. Gallen in Switzerland is a public university and offers undergraduate degrees in Business Administration, Economics, International Affairs, Law and Economics, as well as Legal Studies. Undergraduate degrees take a minimum of three years to complete. The first year serves as a selection and orientation period. Almost all first-year students complete the same set of classes, with few exceptions.[4] Academic performance by the end of

when the students are low-ability students themselves. These results are surprising at first, yet positive effects of low-ability peers have been shown in other contexts (e.g., Bietenbeck, forthcoming).

[4]Exceptions include: Students with non-German mother tongue who choose to complete all first-year courses within two years, and students majoring in Legal Studies. For the latter group, two out of the nine

the first year determines whether the students are admitted to the second year. On average, 66% of students pass the first year in their first attempt. The remaining students either drop out beforehand or fail the first-year exams.[5] After the first year, students choose their major.

All undergraduate degrees start with a mandatory orientation (or "freshmen") week. This week familiarizes the students with the university's infrastructure (e.g., library and online tools) and facilitates contacts among the students. The students are divided into teams at the beginning of the week. Each team consists of 16 students on average, with between 56 and 60 teams per cohort.[6] During the week, students spend approximately 60 hours in their groups, with about 75% of the time dedicated explicitly to team activities (see Appendix Figure A.1). A case study competition between the groups forms the core team activity. The assignment to the groups is fixed, and the students are not allowed to change their group in the course of the week.[7]

The assignment mechanism to freshmen groups is a stratified quasi-random assignment mechanism. The two stratifying variables are gender and the admission requirement that applies to the respective student (some of the students needs to pass an exam in order to be admitted, whereas others are unconditionally admitted).[8] To randomize students into groups, the university administration first fixes the number of groups and then implements the following mechanism. First, students are divided into four strata according to their gender and admission requirement (exam vs. no exam). Second, within each stratum, the students are ordered according to their surnames. Third, student 1 of stratum 1 is placed into the first group, student 2 of stratum 1 into the second group, and so forth, until the stratum is empty. Then, the process starts again: Student 1 of stratum 2 is placed into the first group, student 2 of stratum 2 into the second group, and so forth. This mechanism is repeated for all four strata. Afterwards, an administrator redistributes some students across groups in

mandatory first-year courses differ. Both of these groups combined account for 13% of freshmen.

[5]Students who fail have the option to repeat all first year courses in order to be admitted to the second year.

[6]Group sizes and the number of groups can vary between cohorts. Moreover, group sizes can vary within cohorts because of different room sizes that are available for the groups.

[7]Each group has two group tutors who guide the students through the group activities; the tutors also ensure that the students are in their original groups.

[8]To be precise, all individuals with a non-Swiss high school diploma and a non-Swiss nationality have to pass an admission exam; all other students are unconditionally admitted.

order to match the group sizes to available room sizes. This redistribution is unsystematic.

Since the randomization draws upon the students' surnames, one might object that the process is not fully random—however, the mechanism is designed such that students with similar first syllables of their surnames are likely to end up in different groups. Moreover, in Section 4.3 I show that the resulting groups are balanced in terms of observable student characteristics, conditional on the two stratifying variables.

2.2 Impact of the first week on social ties

If first-week interactions did not impact social ties, they would probably have little effect on behavior in the long run; therefore, this section provides descriptive evidence on the importance of first-week interactions for the formation of friendships and study partnerships. The idea that short interventions matter for longer-term social ties is grounded in results from psychological studies. These studies shows that social contacts are particularly malleable during the first week of college, and random encounters during this period influence social ties that persist for a year or even longer (Back et al., 2008).

To study the importance of first-week interactions for social tie formation, I draw upon two data sources:[9] First, together with the university administration, I ran an online survey of all undergraduate students of cohorts 2008-2010 (2,124 students) who were still enrolled of the time of the survey (May 2012). The survey contained a question on how many of a students' five best friends or study partners were teammates from his/her orientation week; the survey had a response rate of 18%.[10] Second, for the two most recent cohorts (2011-2012), I track how the students selected themselves into one out of 36 tutorial groups, which accompanied the lecture in a mandatory Economics course. In these tutorial groups, the students review the course material together with an instructor. The groups meet every other week, starting in the third week of the academic year.[11] I use attendance data from the

[9] An ideal data set would track the students' social network, including friendships, study partnerships, and acquaintances, and link this data to the administrative sources on student performance. Such data, however, was impossible to collect, both because of data protection demands imposed by the university, and because some students had graduated at the time of the study.

[10] The full questionnaire is available upon request from the author.

[11] All sections are identical in content and take place on the same weekday, but they take place at different

first tutorial—which takes place two weeks after the freshmen week— to detect clustering of teammates in the same tutorial group.

[Figures 1 and 2 about here]

According to the survey data, teammates are overrepresented among a student's five best friends even several months or years after the freshmen week (Figure 1). In the sample of 380 survey respondents, the average number of friends from the same freshmen team among a student's five best friends amounts to 0.68. By contrast, if students chose their friends independently of their team assignment, this fraction would only amount to only 0.08. To obtain this number, I simulate 1,000 random allocations of the five best friends under the assumption that the freshmen week has no impact on friendship formation, and then average across all 1,000 allocations.[12] The difference between the survey and the simulation in the average number of friends who are teammates amounts to 0.60 and is statistically significant ($p < 0.001$).

Since the survey sample is non-representative, I conduct a similar exercise based on selection into tutorial groups. This data covers around 79% of all freshmen students of the cohorts 2011 and 2012. The data on student participation in tutorial groups reveals a similar picture (Figure 2): teammates are overrepresented among members of the same tutorial group. To summarize this, I depict the distribution of teammates (excluding the student himself) among the participants of a student's tutorial group. I also create 1,000 counterfactual allocations of the students across tutorial groups, holding the size of the tutorial groups fixed. I then compare the original distribution to the simulated distribution. The average number of teammates in a student's tutorial group amounts to 0.78 in the original data, but to only 0.39 across all 1,000 permutations. The difference, which amounts to 0.39, is statistically significant ($p < 0.001$).

Taken together, the two analyses present a consistent picture of the persistence of first-

times and have different instructors. In the data, the average number of participants in a tutorial group is 14, with a minimum of six and a maximum of 21 students.

[12]In the simulation, like in the survey, I allow for non-reciprocal friendships. Moreover, I assume that the pool of potential friends is restricted to a student's own cohort, which is more conservative than allowing for friendships across cohorts.

week friendships. Freshmen teammates do not only bond during the first week, but also deepen their bond as they participate in the same activities (e.g., they attend tutorial groups together). At the same time, not all teammates matter. The majority of student remains connected to at most one teammate, if at all. A large number of students still forms their most important friendships and study partnerships outside the freshmen week.

3 Data and descriptive statistics

3.1 Sample

The dataset consists of administrative records for 8,073 freshmen who started their undergraduate degree between 2003 and 2012.[13] Background (pre-treatment) characteristics as well as outcomes are computed from enrollment and grade records. Freshmen group assignments can be matched to these records based on a student identifier. Only a few first-year students had to be deleted from the sample: Students who could not be identified in the freshmen group file, mainly because they did not participate in the freshmen week, as well as students who participated in self-selected freshmen groups.[14] Consisting of 97% of freshmen, the sample is representative for the undergraduate student body.

3.2 Background characteristics

Available pre-treatment characteristics, measured before entry, come from enrolment records. These include: gender, age, nationality (Swiss, German/Austrian, other), mother tongue (German vs. non-German), country of high school degree (Swiss vs. non-Swiss), and state of high school degree for students with a Swiss high school degree (in Switzerland, the states are called "cantons", and Switzerland consists of 26 cantons in total). I use the canton where a student obtained his/her high school degree to proxy for the student's region of origin.

[13] Freshmen group data are unavailable for the year 2005.

[14] Students who had to serve in the army at the time of the freshmen week formed a special group, and up to 3 groups per semester are groups with special tasks ("media groups") for which students could sign up beforehand.

[Table 1 about here]

Panel A of Table 1 presents descriptive statistics of the student body. The majority of students are male (68 percent), and most of the students have a Swiss citizenship (76 percent). 77 percent of the students are from Switzerland; within Switzerland, the students come from culturally different regions. To define regions, I group the regions of origin both by distance to the university and by cultural similarity (Panel A of Table 1); cultural similarity within Switzerland is defined by the official language of the canton.[15] 25 percent of the students are from regions that are in commuting distance to the university. Another 36 percent come from culturally similar cantons, characterized by having German as the only official language in the canton. A smaller fraction comes from culturally different cantons (17 percent). 5 percent of the students with a Swiss citizenship completed their high school diploma outside of Switzerland. 18 percent of the students have neither a Swiss citizenship nor a Swiss high school degree. These students have to pass an admission test in order to be admitted the university. The admission test variable is one of the stratifying variables in the randomization (in addition to gender).

3.3 Treatment: Peer ability

This paper studies ability peer effects based on an imputed measure of ability, similar to the predicted GPA measure used by Carrell et al. (2013). I use imputed ability because the data do not contain any other ability variable that is available for all cohorts and students. To impute a student's ability, I predict each student's first-year GPA using the coefficients from a leave-own-year-out regression of first-year GPA on student background characteristics (see Appendix Table A.2). I standardize the imputed ability measure to have a mean of zero and a standard deviation of one in each cohort. To test the validity of the measure, I compare it to an independent measure of ability, which is available for the cohorts 2011/12. The independent measure is based on a version of Frederick's "Cognitive Reflection Test (CRT)" (Frederick, 2005) and was collected in the course of a classroom experiment (see Schulz et al., 2018,

[15] The official languages are: German, French, Italian, Rumantsch; a canton can have more than one official language. Language as a proxy for culture is used in several studies on Switzerland (e.g., Eugster et al., 2011).

for details of the data collection). The correlation between imputed ability and CRT is 0.27 (p-value < 0.0001, based on 1,627 observations). Moreover, the ability measure is predictive for a range of performance outcomes at college (see Appendix Table A.3).

As the main treatment variable, I use the leave-own-out mean of peers' imputed ability for each student. This variable ranges between -0.92 and 0.75, with a standard deviation of 0.22, which is the variation I exploit to identify ability peer effects.

3.4 Outcomes

The set of outcome variables (Table 2) contains information on the students' performance for up to four years after the students started their undergraduate degree. Moreover, for a subset of students, the data contain survey and experimental measures of confidence.

[Table 2 about here]

A first set of variables represents students' grades (Table 2 Panel A). First-year GPA is computed for all students who have a valid record for at least one compulsory first-year course. Moreover, I study students' performance in a number of compulsory first-semester courses (Mathematics, Economics, Business Administration, Legal Studies). All grades are standardized at the cohort level.

A set of binary variables (Table 2, Panel B) captures dropout behavior and performance during the first year. I use first-semester completion (i.e. completion of all mandatory first-semester courses) as an indicator of initial persistence; 9 percent of the students drop out already within the first few months (between October and February of the academic year). I also include variables indicating whether the students pass the first semester, whether they complete all mandatory courses during the second semester, and whether they pass the second semester.[16] In total, only 66 percent of all freshmen pass the first academic year in their first attempt. These students are automatically promoted to the second year. Many students who do not pass, however, repeat the first year, such that in total 79 percent of all students

[16] Whether the students pass or fail depends on a weighted average of their course grades.

manage to enter the second year eventually.[17]

Major choice takes place after the first year (Table 2, Panel C). Among all majors, Business Administration is the most popular major (66 percent of the students enroll into this major, followed by Economics (16 percent), International Affairs (13 percent), and Law majors (Legal studies or Law & Economics, 12 percent in total). These numbers add up to more than 100 percent, because some of the students pursue double majors (4 percent), and some of the students switch their major at some point during their undergraduate degree (4 percent). Major choice is coded as missing for students who do not start the second year.

The graduation outcomes (Table 2, Panel D) indicate whether the students graduate on time, i.e. within 3 years (28 percent of students), and whether the students graduate within 4 years (82 percent). Furthermore, for those students who graduate within 4 years, I obtain their final GPA.[18] These outcomes are only available for cohorts 2003-2009 because of censoring. I code the graduation outcomes as missing for all students who do not enter the second year.

A final set of outcomes captures the students' confidence in their own academic ability. The confidence measures were elicited two weeks after the freshmen week for cohorts 2011 and 2012 in a pen-and-paper survey and experiment. The questionnaire asked the students to predict their performance rank at the end of the first year, as well as their probability of passing the first year. Separately, the experiment elicited overconfidence based on the students' self-assessment of their performance in a guessing/knowledge task. All measures were standardized to have a mean of 0 and a standard deviation of 1. Appendix Figure A.2 contains details on these variables and their elicitation.

[17]The performance and persistence variables are highly correlated (see also Tafreschi and Thiemann, 2016, who use the same administrative data from St. Gallen).

[18]The final GPA is standardized at the level of the graduation cohort.

4 Empirical approach

4.1 Empirical model

Following Manski (1993), I implement a linear-in-means model of peer effects as the main model specification. Consider a student's test score, which depends on both individual and group characteristics:

$$Y_{igc} = \alpha + \beta x_i + \gamma \overline{x}_{-ig} + \xi s_g + D_c'\delta + W_i'\rho + \epsilon_{igc}, \tag{1}$$

where Y_{igc} is the GPA of student i in cohort c who is a member of freshmen group g, x_i is student $i's$ ability, and $\overline{x}_{-ig}$ represents the average ability of all peers in student $i's$ group, but excluding student i ("leave-own-out mean"). s_g represents the size of group g, D_c is a vector of cohort dummies, W_i is a vector of additional individual-level controls, including the stratifying variables used at randomization (dummy for gender and admission test). ϵ_{igc} is an idiosyncratic error term. In the preferred specifications, I also add individual background characteristics to the vector W_i. These are included to add precision, but do not change the magnitudes of the results.

If the groups are randomly assigned, the coefficient γ identifies the causal effect of group composition on academic achievement, also known as "exogenous" peer effect (Manski, 1993). This model presents a reduced form approach to the estimation of such an exogenous effect (see, for example, Carrell et al., 2013, for a derivation). The conditional expectation of test scores takes the form:

$$\mathbb{E}[Y_{igc}|x_i, \overline{x}_{-ig}, s_g, D_c, W_i] = \alpha + \beta x_i + \gamma \overline{x}_{-ig} + \xi s_g + D_c'\delta + W_i'\rho. \tag{2}$$

The parameter of interest is the average marginal effect of increasing peer ability by one standard deviation. In Model 2, this is γ.

If the outcome variable Y_{igc} is binary (e.g., whether a student drops out during the first semester), a binary response model should be used instead. I assume a logit specification. The conditional probability then takes the form:

$$\mathbb{P}[Y_{igc} = 1 | x_i, \overline{x}_{-ig}, s_g, D_c, W_i] = G(\alpha + \beta x_i + \gamma \overline{x}'_{-ig}\gamma + \xi s_g + D'_c\delta + W'_i\rho), \tag{3}$$

where $G(.)$ is the logistic function.[19] The parameter of interest in the case of a binary outcome is $\theta = E\left[\frac{\partial \mathbb{P}[Y_{igc}=1|X_i,\overline{x}_{-ig},s_g,D_c,W_i]}{\partial \overline{x}_{-ig}}\right]$.

Both empirical studies and theoretical models suggest that students of different ability levels are differentially affected by peer ability (e.g., Carrell et al., 2013, Feld and Zölitz, 2017). Following this literature, I expand the linear-in-means model by allowing for heterogeneous peer effects with respect to a student's own ability. I introduce the variables $high_i$, $middle_i$, and low_i, which take a value of 1 if a student's ability level lies in the top, middle, or bottom tercile of the ability distribution of his/her cohort, respectively, and 0 otherwise. I then allow for different intercepts and slopes for each of the three ability groups:

$$\begin{aligned} Y_{igc} =& \beta_1 high_i + \gamma_1(high_i \times \overline{x}_{-ig}) + \beta_2 middle_i + \gamma_2(middle_i \times \overline{x}_{-ig}) \\ &+ \beta_3 low_i + \gamma_3(low_i \times \overline{x}_{-ig}) + \xi s_g + D'_c\delta + W'_i\rho + \epsilon_{igc}. \end{aligned} \tag{4}$$

The conditional probabilities and average marginal effects can now be computed for each of the three ability groups separately. In the section on robustness checks, I furthermore explore additional specifications, which allow for non-linearity in peer ability in a non-parametric way.[20]

4.2 Estimation and inference

I estimate the model coefficients using OLS for continuous outcomes and using MLE for binary outcomes, and compute the average marginal effects based on these coefficients. The p-values presented throughout are computed using randomization inference (Fisher, 1935).

[19] The results are largely unchanged when using a linear probability model or a probit model instead.

[20] These specifications are based on the fraction of high, middle, and low ability peers in the group, and interactions with own ability.

This inference method is particularly suited to the analysis of randomized experiments. It exploits the assignment protocol directly and provides a transparent way to construct p-values for the null hypothesis of no treatment effect (see Section A.1 for details). In the implementation, I account for clustering at the group level. For recent applications using randomization inference, see for example Carrell and West (2010) and Lim and Meer (2017).

4.3 Test for random assignment

If the randomization has been carried out correctly, a student's characteristics should be unrelated to the composition of the student's peer group, conditional on the stratifying variables used at randomization. To test this assumption, I regress the students' pre-treatment characteristics on peer ability, controlling for the stratifying variables. Appendix Table A.1 presents the results. None of the regressions show a statistically significant relationship between peer ability and the students' pre-treatment characteristics. Thus, the test does not reject the null hypothesis of conditional random assignment.

5 Results

5.1 Main results

Table 3 displays the average marginal effects from linear-in-means regressions of performance outcomes on the average peer ability in a student's peer group.

[Table 3 about here]

The overall effect of average peer ability on students' first-year grades is negative (Table 3, Panel A). An increase of average peer ability by one standard deviation decreases a student's GPA by 7.9 percent of a standard deviation on average. To interpret this magnitude, consider the following thought experiment: If a student was moved from a peer group at the 10th percentile of the peer ability distribution (corresponding to an average peer ability of -0.28) to a peer group in the 90th percentile of the peer ability distribution (corresponding to an average peer ability of 0.27), his/her GPA would decrease by 4.3 percent of a standard deviation on average; this is a statistically significant effect. The effect is similar across courses in both signs and magnitude. While the GPA effect seems rather modest, its magnitude is well in line with prior studies of peer effects on GPA (e.g., Carrell et al. 2009).[21]

Second, higher peer ability increases the chances of initial dropout, but this effect fades out toward the end of the first year. An increase in average peer ability by one standard deviation decreases the chance of completing all first-semester courses by 2.4 percentage points. To continue the above example, if a student was moved from the 10th to the 90th decile of peer ability, his/her probability of early dropout would increase by 1.3 percentage points on average. This effect is large and amounts to 12 percent of the average dropout probability in the sample (9 percent). However, the effect gradually fades out—students with higher peer ability are neither less likely to pass the first year nor less likely to start the second year.

Third, higher peer ability negatively affects the probability of pursuing a double major as well as the probability of pursuing a Business Administration major, which is the most popular major. An increase in average peer ability by one standard deviation decreases the probability of choosing a Business Administration major by 6 percentage points on average. This effect is partly driven by those students who decide not to pursue a double major as a result of higher average peer ability. If a student was moved from the 10th to the 90th decile of the peer ability distribution, his/her probability of choosing a double major would decline by 1.2 percentage points. This amounts to 29 percent of the average probability of pursuing a double major in the sample (4 percent). The effect of peer ability on the less popular majors is small and insignificant.

Finally, peer ability during the first week negatively affects the students' final GPA, but does not affect the time that the students need in order to graduate, conditionally on starting the second year. The GPA at graduation decreases by 15.3 percent of a standard deviation as

[21] In their seminal study carried out at the Air Force Academy, Carrell et al. (2009) find that an increase in peers' verbal scholastic assessment test (SAT) score by one standard deviation on average *increases* freshmen GPA by 0.08 standard deviations. In this study, I find that an increase in peer ability by one standard deviation *decreases* peer ability by about 0.08 standard deviations.

a result of an increase of average peer ability by 1 standard deviation. In the above example, moving a student from the 10th to the 90th percentile of average peer ability increases his/her final GPA by 8.4 percent of a standard deviation. This effect is again modest yet significant. It is important to keep in mind that the results on final GPA come from a selected sample, i.e. from the students who had graduated within four years. Since the probability of entering the second year as well as the graduation probabilities do not respond to changes in the peer group ability, however, the effect of peer ability on final GPA, conditional on graduating, is likely not a (pure) selection effect. I will further explore this issue in the heterogeneity analysis in Section 5.2.

To sum up, the overall effect of peer ability on student performance is negative across a wide range of outcomes, and, for some outcomes, persists throughout the whole undergraduate degree. The grade outcomes are most affected, even until the time of graduation. Beyond increases in initial dropout, the retention and graduation outcomes are not affected. This suggests that high-ability peers induce early dropout primarily in those students who might have dropped out at a later stage in any case. The heterogeneity analysis in Section 5.2 explores this point in more detail. Moreover, higher peer ability deters students from pursuing a double major.

The evidence on negative peer effects might seem surprising at first sight, since one might assume that high-ability students can share their knowledge and positive habits with their peers and thus generate positive spillovers. The negative effect, however, can be interpreted as a "discouragement effect": High-ability students might set unattainable performance expectations—especially for low-ability students—and thus deter other students form exerting effort or completing a course (e.g., Rogers and Feller, 2016). The following section therefore tests whether low-ability students are especially vulnerable to the presence of high-ability peers.

5.2 Heterogeneity

Table 4 presents the results of the heterogeneity analysis, which includes interactions between a students' own ability level and the average ability of the peer group (see Model 4). For brevity, the table and discussion concentrate on five key outcomes which represent the students' trajectory well: first-year GPA as a summary measure of first-year performance; first-semester completion as a measure of initial retention ("no initial dropout"); starting the second year as a measure of longer-term retention, choosing a double major as a measure of ambition, and final GPA as a long-term performance measure.

[Table 4 about here]

First, low-ability students are negatively affected by their peers in terms of their first-year grades, initial retention, and in terms of whether they make it into the second year, whereas both high- and middle-ability students do not experience such negative effects. The effect of peer ability on first-year GPA for the low-ability students is three times as large as the effect in the whole sample. If a student was moved from the 10th to the 90th percentile of peer ability, his/her first-year GPA would decrease by 14 percent of a standard deviation. The same change would affect the probability of initial dropout for low-ability students by 2.7 percent, which amounts to an increase of 16 percent of the baseline—the average initial dropout probability is 17 percent for low-ability students (see Table A.3 for summary statistics of outcomes by ability level). This effect translates almost fully into a lower probability of starting the second year. Thus, the dropout effect does not fade out for the low-ability students. For high- and middle-ability students, all effects on first-year outcomes are small and insignificant.

Second, high-ability students are less likely to pursue a double major when exposed to high-ability peers; neither middle- nor low-ability students experience negative effects on the probability of choosing a double major. For high-ability students, the probability of pursuing a double major drops by 1.3 percent if they moved from a group at the 10th percentile to a group at the 90th percentile of peer ability. This effect is large compared to the baseline; it amounts to 19 percent of the average probability of choosing a double major for high-ability students (7 percent, see Table A.3). I find no significant effects for low- and middle-ability students.

Finally, the effect on final GPA is negative for high-, middle-, and low-ability students, but only statistically significant for middle-ability students. For middle ability students, a shift from a group at the 10th percentile of peer ability to a group at the 90th percentile of peer ability decreases final GPA by 0.14 standard deviations. The fact that the low-ability students are less affected than the middle-ability students can be a compositional effect: out of those low-ability students who react to peer ability, some have already dropped out during the first year.

To sum up, the initial effects of high-ability peers are concentrated among low-ability students. Among those students, grades decrease as a response to high-ability peers, which translates into higher dropout rates. By contrast, the effects that occur during later stages—major choice and final GPA—are rather concentrated among high- and middle-ability students.

5.3 Channels

This section investigates "discouragement effects" as potential channels for negative peer effects. A discouragement effect is at play if students become intimidated once they are exposed to high-ability peers. This can negatively affect the students' effort and commitment and ultimately lower their academic achievement. I use behavioral measures of confidence, elicited shortly after the freshmen week for two cohorts (2011 and 2012), to provide descriptive evidence for this hypothesis. Specifically, I test whether peer ability is negatively correlated with a student's confidence after the intervention finished.

Table 5 presents results of regressions of three confidence measures on the average peer ability in the group. The results show a negative relationship between peer ability and all confidence measures. Signs and magnitudes of all three confidence measures are similar to each other. Perceived rank as well as perceived passing probability are weakly significantly associated with peer ability, whereas the overconfidence measure from the experimental task is not significantly associated with peer ability.

[Table 5 about here]

The negative correlation between peer ability and confidence, however, needs to be interpreted with caution. First, the results are only available for two cohorts, limiting the sample size and the scope for further analysis (e.g., analysis of effect heterogeneity). Second, the participation in the survey was voluntary, which might lead to a selected sample and cause potential selection bias in the estimates.[22] The main results in the survey sample are similar to the results in the full sample (see Tables 3 and 5). The effects on first-year GPA and on initial dropout probabilities, however, are larger in the survey sample, because these effects are stronger in the two most recent cohorts, compared to the full sample. To establish a causal relationship between peer ability and confidence remains a subject of further research.

6 Functional form and robustness checks

6.1 Functional form

Allowing for non-linearity in peer ability. The linear-in-means model does not capture non-linearities in the treatment effect. I therefore test a specification that uses the fraction of high- and low-ability peers instead of average peer ability as a treatment variable. Appendix Table A.4 displays the results of the non-linearity analysis for the five key variables. In addition to the coefficients on the fraction of high- and low-ability peers, the table presents the difference of the coefficients and their p-value. The fraction of middle-ability students serve as the reference category.

The results do not display a clear non-linear pattern. While increases in the fraction of high-ability peers have a negative impact on the outcomes, increases in low-ability peers

[22] Students with higher average levels of peer ability participate less frequently in the survey (coefficient of -0.135, significant at the 1-percent level). If high-confidence students select out of the survey as a result of being in a higher-ability peer group, this would imply that higher-ability peer groups were left with lower-confidence peers; such a selection pattern would bias the effect downwards, i.e. away from zero.

have a positive impact. The choice of the reference category is important to consider when interpreting the results. Replacing *middle-ability* peers with high-ability peers does not have a significantly negative impact on four out of five outcomes—as denoted by the coefficients on the "fraction high ability peers". By contrast, replacing *low-ability* peers with high ability peers has a strongly negative and significant impact on all five outcomes—as denoted by the coefficient on the "difference fraction high-low". For example, if the share of high-ability students rises by 10 percentage points, and the share of low-ability students declines by 10 percentage points in turn, a student's test scores would drop by 2.6 percent of a standard deviation. In the cross-section, this effect persists until graduation.

In sum, I find no clear evidence of non-linearities; overall, high-ability students have a positive, and low-ability students have a negative impact. The analysis clearly reveals that the negative effects are driven by the fraction of high-ability students.

Combining non-linearity and heterogeneity. Since high-, middle-, and low-ability students react differently to increases in peer ability, I also present a specification that combines non-linearity in the treatment with an analysis of heterogeneity between subgroups. In other words, I test whether the fraction of high-, middle-, and low-ability peers has different effects across students of different ability levels. Appendix Tables A.5 and A.6 display the results.

Overall, the results are similar to the results from the heterogeneity analysis in Section 5.2, but allowing for non-linearities reveals some patterns that the linear-in-means model overlooks. The effects are, again, strongest for low-ability students: When the share of high-ability peers rises and the share of low-ability peers drops by the same amount, the performance and attainment of low-ability students is negatively affected. The effects on first-year GPA, first-semester completion, and entering the second year are highly significant. For middle-ability students, I find a non-linear relationship between peer ability and first-semester completion: both low- and high-ability peers affect the probability of finishing the first semester positively. For high-ability students, the only significant impacts are on the probability of pursuing a double major. Here, the analysis is in line with the linear-in-means model as well. Finally, high-ability peers have a negative impact on final GPA in all subgroups, but only the middle-ability students display a weakly significant effect, and all other groups are not significantly affected.

In sum, the combination of non-linearity and heterogeneity confirms the findings from Section 2, and shows that the linear-in-means model provides a good approximation for most of the outcomes. In some cases, the analysis reveals non-linearities, showing that it can be important to take both non-linearities and heterogeneity into account before considering interventions which change assignments of students to peer groups.

6.2 Robustness

Control variables. Since the administrators randomized the students into groups, the coefficients should not change when adding control variables to the specification. In the main part, my preferred specifications contain control variables because they allow for a more precise estimation of some of the coefficients. Appendix Table A.7 shows that coefficients and significance levels from the main part do not change appreciably when running the regression without control variables.

Alternative samples. Graduation outcomes are only available for the older cohorts (2003-2009), because the younger cohorts had not graduated at the time of the data collection. To check whether the cohorts of 2003-2009 are representative for the whole sample, Appendix Table A.8 replicates the main regressions for these cohorts only. The results for the cohorts of 2003-2009 are in line with the results for the full sample.

Alternative ability categories. To estimate models with effect heterogeneity and non-linearities, I divided the ability distribution into three terciles. I check whether the analysis is robust to alternative definitions of high and low ability and define the high-ability students as the students in the top quartile the ability distribution, and the low-ability students as the students in the bottom quartile of the ability distributions. The results of both the heterogeneity analysis and the analysis of non-linearities are robust to changes in the definition of high- and low-ability students (Appendix Tables A.9 and A.10).

7 Conclusion

This paper studies peer effects from a mandatory introductory week for undergraduate students at a Swiss university. It is thus the first paper to exploit a short intervention in a higher education setting and investigate its implications for social tie formation and subsequent academic outcomes.

As a first result, the present study shows that peer groups based on a one-week intervention can generate persistent peer effects on final GPA, dropout, and major choice. This is in line with a few prior studies that document persistent effect of initial peer groups in university settings. Lyle (2007) finds an effect of peers' attitude toward the army on a student's probability of remaining in the army after six years, therefore showing a long-term impact. Similarly, Carrell et al. (2009) find that positive peer effects of peers' verbal SAT on students' GPA persist at least until the senior year. By contrast, Sacerdote (2001) finds no persistent effects of roommates' academic ability on students' outcomes after the freshmen year.

As a second result, this paper finds negative effects of peer ability on performance, dropout and ambition (choosing a double major). The negative effects on first-year outcomes—initial dropout and performance—are concentrated among the low-ability students. While many studies find positive ability peer effects in university contexts (Booij et al., 2017, Carrell et al., 2009), a number of studies also document negative effects, at least for subgroups. In the context of higher education, Feld and Zölitz (2017) find that low-ability students' course grades are harmed by high-ability peers in a tutorial group. Using data from the US Airforce Academy, Carrell et al. (2013) show that high fractions of high-ability students can harm low-ability students in cases where the groups consist of high- and low-ability students only. In the context of primary education, Antecol et al. (2016) find that increases in average peer ability have a negative impact on students' performance in elementary schools that are in disadvantaged neighborhoods. Thus, so far no consensus exists as to whether peer ability effects should be positive and negative. All studies emphasize that the context of the study matters. In the context of the present paper, the students enter into a highly selective environment: only two-thirds of all students pass the first year in their first attempt. This environment is intimidating in itself for many students, and the presence of high-performing peers might exacerbate the perception of how difficult it is to succeed.

Moreover, I link the findings in this study to a literature on discouragement effects. The results supported by experimental evidence from social psychology. Rogers and Feller (2016) show that students who were randomly exposed to excellent peers when completing a writing task of a course were more likely to quit the course, compared to students who were not exposed to excellent peers. This is in line with the present study: higher peer ability increases the probability of initial dropout. The results also relate to a literature which argues that students are motivated by minimizing the performance distance to other students. If the performance distance appears too large, students might resort to low effort because they cannot keep up (e.g., Tincani, 2018).

Finally, it is important to qualify the finding that high-achievers may generate negative peer effects within their immediate group. At the cohort level, the presence of high-ability students may increase the pace or quality of instruction. It is thus unclear whether the inflow of highly selected students is beneficial overall; but a principal may in any case consider how to integrate high-ability students with their cohort members in order to minimize any adverse spillovers.

References

Altonji, J. G., P. Arcidiacono, and A. Maurel (2016): "The analysis of field choice in college and graduate school: Determinants and wage effects," in *Handbook of the Economics of Education*, Elsevier, vol. 5, 305–396.

Antecol, H., O. Eren, and S. Ozbeklik (2016): "Peer Effects in Disadvantaged Primary Schools Evidence from a Randomized Experiment," *Journal of Human Resources*, 51, 95–132.

Back, M. D., S. C. Schmukle, and B. Egloff (2008): "Becoming Friends by Chance," *Psychological Science*, 19, 439–440.

Bietenbeck, J. (forthcoming): "The Long-Term Impacts of Low-Achieving Childhood Peers: Evidence from Project STAR," *Journal of the European Economic Association*.

Booij, A. S., E. Leuven, and H. Oosterbeek (2017): "Ability Peer Effects in University: Evidence from a Randomized Experiment," *The Review of Economic Studies*, 84, 547–578.

Brenøe, A. A. and U. Zölitz (2018): "Exposure to More Female Peers Widens the Gender Gap in STEM Participation," Working Paper 285, University of Zurich, Department of Economics.

Carrell, S. E., R. L. Fullerton, and J. E. West (2009): "Does Your Cohort Matter? Measuring Peer Effects in College Achievement," *Journal of Labor Economics*, 27, 439–464.

Carrell, S. E., B. I. Sacerdote, and J. E. West (2013): "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," *Econometrica*, 81, 855–882.

Carrell, S. E. and J. E. West (2010): "Does professor quality matter? Evidence from random assignment of students to professors," *Journal of Political Economy*, 118, 409–432.

Epple, D. and R. E. Romano (2011): "Peer Effects in Education: A Survey of the Theory and Evidence," in *Handbook of Social Economics*, ed. by J. Benhabib, A. Bisin, and M. O. Jackson, North-Holland, vol. 1, chap. 20, 1053–1163.

Eugster, B., R. Lalive, A. Steinhauer, and J. Zweimüller (2011): "The Demand for Social Insurance: Does Culture Matter?" *The Economic Journal*, 121, 413–448.

Feld, J. and U. Zölitz (2016): "Majors, Math, & Marks - Gender Peer Effects in University," Manuscript.

——— (2017): "Understanding Peer Effects: On the Nature, Estimation and Channels of Peer Effects," *Journal of Labor Economics*, 35, 387–428.

Fisher, R. A. (1935): *The Design of Experiments*, Oliver and Boyd.

Frederick, S. (2005): "Cognitive Reflection and Decision Making," *The Journal of Economic Perspectives*, 19, 25–42.

Hoxby, C. M. and G. Weingarth (2005): "Taking race out of the equation: School reassignment and the structure of peer effects," Manuscript.

Lim, J. and J. Meer (2017): "The Impact of Teacher-Student Gender Matches: Random Assignment Evidence from South Korea," *Journal of Human Resources*, 52, 979–997.

Lyle, D. S. (2007): "Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point," *The Review of Economics and Statistics*, 89, 289–299.

——— (2009): "The Effects of Peer Group Heterogeneity on the Production of Human Capital at West Point," *American Economic Journal: Applied Economics*, 1, 69–84.

Manski, C. F. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60, 531–42.

Oosterbeek, H. and R. van Ewijk (2014): "Gender peer effects in university: Evidence from a randomized experiment," *Economics of Education Review*, 38, 51–63.

Rogers, T. and A. Feller (2016): "Discouraged by Peer Excellence: Exposure to Exemplary Peer Performance Causes Quitting," *Psychological Science*, 27, 365–374.

Sacerdote, B. (2001): "Peer Effects With Random Assignment: Results For Dartmouth Roommates," *The Quarterly Journal of Economics*, 116, 681–704.

——— (2011): "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" in *Handbook of the Economics of Education*, ed. by E. Hanushek, S. Machin, and L. Woessmann, Elsevier, vol. 3, chap. 4, 249–277.

Schulz, J., P. Thiemann, and C. Thöni (2018): "Nudging Generosity: Choice Architecture and Cognitive Factors in Charitable Giving," *Journal of Behavioral and Experimental Economics*, 74, 139–145.

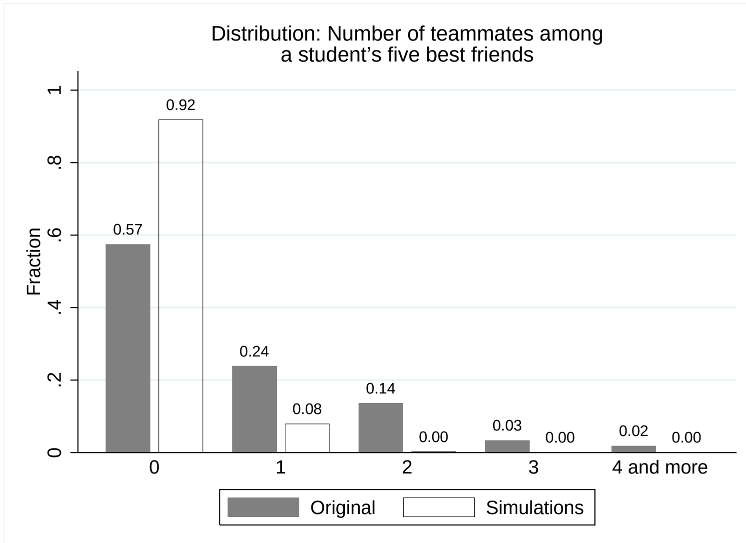
Stinebrickner, R. and T. R. Stinebrickner (2006): "What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds," *Journal of Public Economics*, 90, 1435–1454.

Tafreschi, D. and P. Thiemann (2016): "Doing it Twice, Getting it Right? The Effect of Grade Retention and Course Repetition in Higher Education," *Economics of Education Review*, 55, 198–219.

Tincani, M. M. (2018): "Heterogeneous Peer Effects in the Classroom," Manuscript.

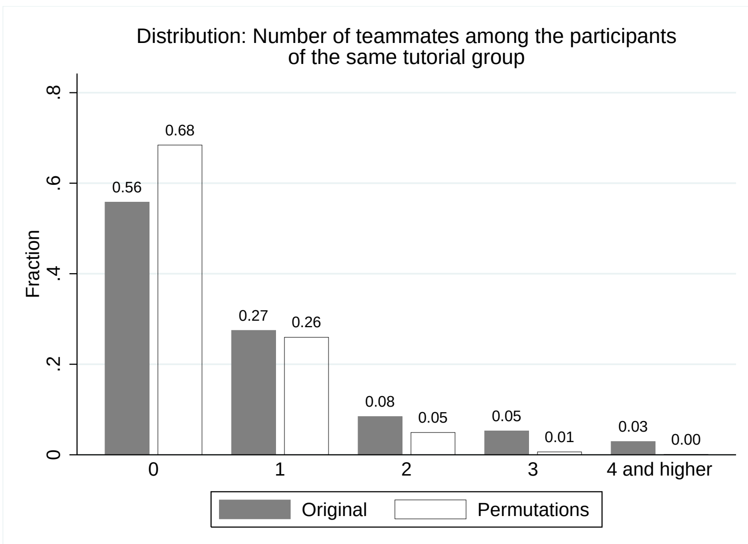
Figures and tables

Figure 1: Friendship formation



Note: The dark bars show the actual distribution of the number of teammates from the freshmen week among a student's five best friends based on survey data. The white bars show the same distribution, based on 1,000 simulations under the assumption that students form their friendships independent of freshmen teams. The survey data contain a subsample of students from cohorts 2008-2010 (380 observations). The simulation is based on the full sample of cohorts 2008-2010 (3,024 observations). For details on the simulation, see Section 2.2. The average number of teammates among the five best friends is 0.68 in the survey data, and 0.08 on average across 1,000 simulations. The difference is significant ($p < 0.001$).

Figure 2: Selection into the same tutorial groups



Note: The dark bars show the actual distribution of the number of teammates from the freshmen week in a student's tutorial group, based on participation data. The white bars show the same distribution, based on 1,000 simulations under the assumption that students participate in tutorial groups independent of other members of their freshmen team. The data contain all tutorial-group participants of cohorts 2011/12 (1,666 students). For details on the simulation, see Section 2.2. The average number of teammates among all participants in a student's tutorial group is 0.78 in the original data and 0.39 on average across 1,000 permutations. The difference is significant ($p < 0.001$).

Table 1: Descriptive statistics: Student background and treatment

	Mean	SD	Min.	Max.	Obs.
A. Student background					
Female (D)	32%	-	0	1	8,073
Age	20.1	1.9	16	48	8,073
Non-Swiss nationality (“foreign”) (D)	24%	-	0	1	8,073
Non-Swiss mother tongue (D)	11%	-	0	1	8,073
Region of origin (region of high school degree)					
Close to the university (German-speaking) (D)	25%	-	0	1	8,073
Remaining German-speaking cantons (D)	36%	-	0	1	8,073
Mixed-language cantons (D)	11%	-	0	1	8,073
Non-German speaking cantons (D)	6%	-	0	1	8,073
Swiss nationality, HS foreign (D)	5%	-	0	1	8,073
Foreign, HS foreign (D) (“Admission test”)	18%	-	0	1	8,073
Ability (imputed)	0.00	1.00	-5.91	2.76	8,073
B. Treatment					
Peer ability	0.00	0.22	-0.92	0.75	8,073
C. Group variable					
Group size	16	3.0	7	22	8,073

Note: The table presents descriptive statistics of the estimation sample (cohorts 2003-2012), based on administrative records.

Table 2: Descriptive statistics: Student outcomes

	Mean	SD	Min.	Max.	Obs.
A. First-year grades					
First year grade-point average (GPA)	0.00	1.00	-4.19	1.92	7,953
Math grade	0.01	1.00	-2.97	1.72	7,405
Economics grade	0.01	1.00	-3.70	1.96	7,848
Business administration grade	0.00	1.00	-4.31	2.42	7,909
Legal studies grade	0.00	1.00	-3.79	2.22	7,703
B. First-year retention					
First semester: all courses completed (D)	91%	-	0	1	8,073
First semester: passed (D)	79%	-	0	1	8,073
Second semster: all courses completed (D)	74%	-	0	1	8,073
Second semester: passed (D)	66%	-	0	1	8,073
Second year started (D)	79%	-	0	1	8,073
C. Major choice					
Business administration (D)	66%	-	0	1	6,343
Economics (D)	16%	-	0	1	6,343
International Affairs (D)	13%	-	0	1	6,343
Law & Economics/Legal Studies (D)	12%	-	0	1	6,343
Double Major (D)	4%	-	0	1	6,343
D. Graduation					
Completed degree within 3 years (on time) (D)	28%	-	0	1	3,953
Completed degree within 4 years (D)	82%	-	0	1	3,953
Final GPA	0.10	0.98	-2.44	2.84	3,241
E. Confidence					
Perceived rank	0.00	1.00	-3.61	2.30	1,657
Perceived passing probability	0.00	1.00	-4.47	3.55	1,654
Over-confidence in experimental task	0.01	1.00	-2.67	3.04	1,368

Note: The table presents descriptive statistics of the estimation sample (cohorts 2003-2012), based on administrative records. First-year GPA (Panel A) is computed for all students who have a valid grade record for at least one first-year course. First-year grades for the compulsory courses math, economics, business administration, and legal studies (Panel A) are only reported for those students with a valid record in the respective course. Grades are standardized by cohort. Data on major choice (Panel C) is only computed for individuals who enter the second year. Data on graduation outcomes (Panel D) is only available for cohorts 2003-2009 because of censoring. On-time graduation is only computed for individuals who enter the second year, and final GPA is only computed for individuals who graduate within four years. Survey data on confidence (Panel E) is only available for cohorts 2011-2012 and for individuals who participated in a pen-and-paper survey/experiment.

Table 3: Effects of peer ability: Linear-in-means model

	(1)	(2)	(3)	(4)	(5)
	Panel A. First-year grades				
	GPA	Math	Economics	Business	Legal Studies
Peer ability	-0.079**	-0.090**	-0.070*	-0.097**	-0.072*
p-value	(0.033)	(0.021)	(0.056)	(0.018)	(0.050)
N	7,953	7,405	7,848	7,909	7,703
	Panel B. First-year retention				
	1st semester		2nd semester		2nd year
	completed	passed	completed	passed	started
Peer ability	-0.024**	-0.030*	-0.017	0.000	-0.009
p-value	(0.038)	(0.062)	(0.211)	(0.525)	(0.290)
N	8,073	8,073	8,073	8,073	8,073
	Panel C. Major choice				
	Business	Economics	International Affairs	Law	Double major
Peer ability	-0.058**	0.016	0.009	0.004	-0.021**
p-value	(0.021)	(0.251)	(0.334)	(0.432)	(0.021)
N	6,343	6,343	6,343	6,343	6,343
	Panel D. Graduation				
	3 years (on time)	4 years	Final GPA		
Peer ability	-0.002	-0.026	-0.153**		
p-value	(0.509)	(0.131)	(0.013)		
N	3,953	3,953	3,241		

Note: The table presents average marginal effects from regressions of student outcomes on average peer ability, using linear models for continuous outcomes, and logit models for binary outcomes (cohorts 2003-2012). All regressions include the stratifying variables (gender, admission test, cohort dummies), and a set of additional controls (own ability, region-of-origin dummies, mother tongue, age, group size). P-values are computed using randomization inference (1,000 draws from the randomization distribution). Graduation outcomes are only available for cohorts 2003-2012.
*p < 0.10, **p < 0.05, ***p < 0.01.

Table 4: Effect heterogeneity by own ability

	(1) 1st year GPA	(2) 1st semester completed	(3) 2nd year started	(5) Double major	(6) Final GPA
Peer ability					
× high	-0.033	-0.013	0.008	-0.024**	-0.101
	(0.264)	(0.342)	(0.415)	(0.016)	(0.143)
Peer ability					
× middle	0.047	0.025	0.025	-0.014	-0.270**
	(0.278)	(0.201)	(0.261)	(0.381)	(0.020)
Peer ability					
× low	-0.262***	-0.049***	-0.045*	-0.012	-0.100
	(0.004)	(0.003)	(0.060)	(0.348)	(0.202)
N	7,953	8,073	8,073	6,343	3,241

Note: The table presents average marginal effects from regressions of student outcomes on average peer ability, interacted with a students' own ability (classified as high, middle, and low). The regressions are based on linear models for continuous outcomes and logit models for binary outcomes. All regressions include the stratifying variables (gender, admission test, cohort dummies), and a set of additional controls (own ability, region-of-origin dummies, mother tongue, age, group size). P-values are computed using randomization inference (1,000 draws from the randomization distribution). Final GPA is only available for cohorts 2003-2012.
*p $<$ 0.10, **p $<$ 0.05, ***p $<$ 0.01.

Table 5: Regressions of confidence measures on peer ability

	(1)	(2)	(3)	(4)
	Panel A. Confidence in academic ability			
	Perceived rank	Perceived passing probability	Over-confidence	
Peer ability	-0.138*	-0.167*	-0.133	
p-value	(0.069)	(0.062)	(0.137)	
N	1,657	1,654	1,368	
	Panel B. Performance outcomes			
	1st year GPA	1st semester completed	2nd year started	Double major
Peer ability	-0.187**	-0.057**	-0.049	0.022
p-value	(0.030)	(0.031)	(0.114)	(0.198)
N	1,660	1,674	1,674	1,356

Note: The tables presents average marginal effects from regressions of student outcomes on peer ability, using linear models for continuous outcomes and logit models for binary outcomes. The confidence measures were collected in a pen-and-paper survey two weeks after the orientation week. The sample contains all individuals in the cohorts who participated in the survey and answered at least one of the questions on confidence. The participation rate in the survey was 79%, and individuals with higher peer ability were less likely to participate (coefficient of -0.12 with a p-value of 0.004). All regressions include the stratifying variables (gender, cohort dummies, admission test), and a set of additional controls (own ability, region-of-origin dummies, mother tongue, age, group size). P-values are computed using randomization inference (1,000 draws from the randomization distribution).

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

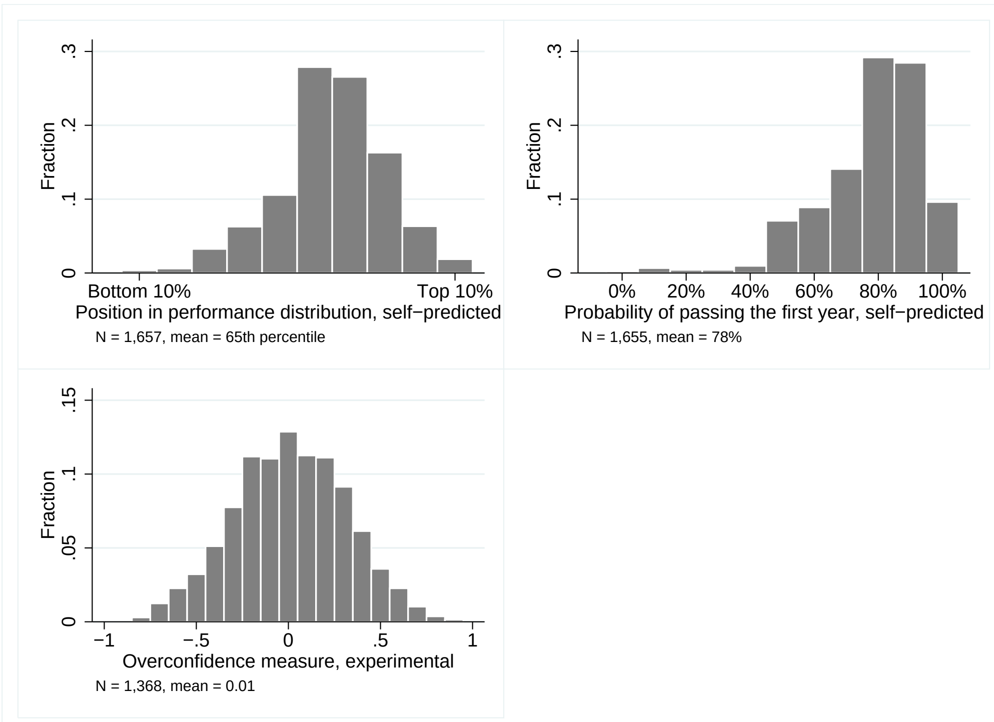
Appendix

Figure A.1: Schedule of the orientation week for a typical group

<table>
<tr><th></th><th>Monday</th><th>Tuesday</th><th>Wednesday</th><th>Thursday</th><th>Friday</th></tr>
<tr><td rowspan="2">8 a.m.</td><td rowspan="4">Welcome (auditorium)</td><td rowspan="2">Input talk for the case study</td><td rowspan="2">Input talk for the case study</td><td rowspan="10">Case study</td><td rowspan="2">Competition: first round</td></tr>
<tr></tr>
<tr><td rowspan="2">9 a.m.</td><td rowspan="2">Introduction to the library</td><td rowspan="6">Case study</td><td rowspan="1">Break</td></tr>
<tr><td rowspan="11">Competition: second round in auditorium</td></tr>
<tr><td rowspan="2">10 a.m.</td><td rowspan="2">Team building</td><td rowspan="2">Campus walk</td></tr>
<tr></tr>
<tr><td rowspan="2">11 a.m.</td><td rowspan="2">Lunch</td><td rowspan="1">Case study</td></tr>
<tr><td rowspan="1">Lunch</td></tr>
<tr><td rowspan="2">12 p.m.</td><td rowspan="2">Case study</td><td rowspan="8">Case study</td><td rowspan="1">Lunch</td></tr>
<tr><td rowspan="11">Case study</td></tr>
<tr><td rowspan="2">1 p.m.</td><td rowspan="6">Introduction to university infrastructure</td><td rowspan="1">Lunch</td></tr>
<tr><td rowspan="9">Case study</td></tr>
<tr><td rowspan="2">2 p.m.</td></tr>
<tr></tr>
<tr><td rowspan="2">3 p.m.</td><td rowspan="14">Free afternoon</td></tr>
<tr></tr>
<tr><td rowspan="2">4 p.m.</td><td rowspan="4">Case Study</td><td rowspan="4">Introduction to exam procedures</td></tr>
<tr></tr>
<tr><td rowspan="2">5 p.m.</td></tr>
<tr></tr>
<tr><td rowspan="2">6 p.m.</td><td rowspan="2">Dinner</td><td rowspan="2">Dinner</td><td rowspan="2">Dinner</td><td rowspan="2">Dinner</td></tr>
<tr></tr>
<tr><td rowspan="2">7 p.m.</td><td rowspan="8">Team evening</td><td rowspan="2">Alumni event</td><td rowspan="8">Team evening</td><td rowspan="8">Student club presentations</td></tr>
<tr></tr>
<tr><td rowspan="2">8 p.m.</td><td rowspan="6"></td></tr>
<tr></tr>
<tr><td rowspan="2">9 p.m.</td></tr>
<tr></tr>
<tr><td rowspan="2">10 p.m.</td><td rowspan="2">Student party</td></tr>
<tr></tr>
</table>

Note: Orientation week schedule of a typical group. Dark grey areas indicate time slots spent only in the assigned group, light grey areas indicate time slots spent in assigned groups, but possibly together with other groups, white areas indicate time slots spent not necessarily in assigned groups.

Figure A.2: Histograms of confidence measures



Note: The figure shows histograms of the confidence measures collected in a pen-and-paper survey/experiment for cohorts 2011-2012. To elicit confidence in one's own performance (top left), we asked the students: "Suppose that you completed all exams in the first year. At the end of the first year, where do you think you will rank among all other students who have completed all first-year exams?" The students had to answer this question on a scale of 1-10. To elicit self-predictions of the passing probability (top right), we asked the students to rate their passing probability on a scale of 1-10. In the analysis, we correct this measure for the students' guess of the mean passing rate. To elicit overconfidence (bottom left), we first asked the students to complete a guessing task. After completion, we asked the students to assess their performance on this task on a scale of 1-10. The overconfidence task was incentivized. We standardized the measure to have a mean of 0 and a standard deviation of 1.

Table A.1: Test for random assignment

Dependent variable:	(1) Coeff.	(2) Median	(3) p-value
Ability	-0.033	-0.012	(0.326)
Age (years)	0.092	0.020	(0.232)
Non-Swiss nationaliy	0.002	0.003	(0.525)
Non-German mother tongue	-0.004	0.005	(0.329)
Region of origin (High school degree)			
close to the university	0.003	-0.001	(0.436)
mixed-language canton	0.013	0.000	(0.249)
non-German speaking canton	0.008	0.002	(0.342)
HS degree non-Swiss	0.000	0.002	(0.559)
Number of students		8,073	
Number of groups		526	

Note: The table presents average marginal effects from regressions of average peer ability on student background characteristics. Each row presents the results of a single regression, with a different dependent variable per row. The regressions control for the stratifying variables (gender, admission test, and cohort dummies) as well as for group size. P-values (Column 3) are in parentheses and computed using randomization inference (1,000 draws from the randomization distribution). The table also shows the median of the randomization distribution (Column 2).
*p $<$ 0.10, **p $<$ 0.05, ***p $<$ 0.01.

Table A.2: Imputation of ability

Dependent variable: GPA	
Gender	
Female	-0.111***
	(0.024)
Region of origin (Ref: German-speaking cantons)	
Close to the university	-0.050*
	(0.028)
Mixed-language canton	-0.047
	(0.038)
Non-German speaking cantons	-0.275***
	(0.066)
Swiss, high school degree foreign	-0.373***
	(0.054)
Foreign, high school degree foreign	0.755***
	(0.052)
Further characteristics	
Non-German mother tongue	-0.277***
	(0.049)
Non-Swiss nationality	-0.414***
	(0.045)
Age (years)	-0.160***
	(0.031)
Age (squared)	0.002***
	(0.001)
Constant	2.563***
	(0.391)
Adjusted R-squared	0.080
Obs.	7,953

Note: The table presents coefficients from OLS regressions of first-year GPA on student characteristics.
*p < 0.10, **p < 0.05, ***p < 0.01.

Table A.3: Descriptive statistics: Outcomes by ability level

	Mean high ability	Mean middle ability	Mean low ability
A. First-year grades			
First year grade-point average (GPA)	0.28	0.01	-0.32
Math grade	0.25	-0.05	-0.25
Economics grade	0.31	-0.04	-0.32
Business administration grade	0.25	0.05	-0.34
Legal studies grade	0.21	0.01	-0.26
B. First-year retention			
First semester: all courses completed (D)	96%	94%	83%
First semester: passed (D)	88%	81%	67%
Second semster: all courses completed (D)	83%	78%	61%
Second semester: passed (D)	77%	68%	51%
Second year started	87%	80%	67%
C. Major choice			
Business administration (D)	71%	61%	64%
Economics (D)	20%	12%	15%
International Affairs (D)	12%	14%	15%
Law & Economics/Legal Studies (D)	9%	17%	13%
Double major (D)	7%	1%	1%
D. Graduation			
Completed degree within 3 years (on time) (D)	27%	27%	31%
Completed degree within 4 years (D)	84%	82%	80%
Final GPA	0.39	-0.06	-0.17
E. Confidence			
Perceived rank	0.44	-0.25	-0.33
Perceived passing probability	0.26	-0.16	-0.19
Over-confidence in experimental task	0.07	-0.05	-0.02

Note: The tables presents average outcomes for students with high, middle, and low predicted GPA (divided by terciles of the predicted GPA distribution).

Table A.4: Effects of peer ability: Non-linear model

	(1) 1st year GPA	(2) 1st semester completed	(3) 2nd year started	(4) Double major	(5) Final GPA
Fraction high ability peers	-0.145*	0.002	-0.030	-0.007	-0.142
p-value	0.096	0.459	0.228	(0.376)	(0.192)
Fraction low ability peers	0.110	0.044*	0.010	0.025	0.125
p-value	0.121	0.072	0.421	(0.145)	(0.196)
Difference fraction high - low	-0.255**	-0.042*	-0.040	-0.032*	-0.267**
p-value	0.011	0.087	0.187	0.084	0.047
N	7,953	8,073	8,073	6,343	3,241

Note: The table presents average marginal effects from regressions of student outcomes on the fraction of high and low ability peers, using linear models for continuous outcomes, and logit models for binary outcomes. Middle-ability peers are in the reference category. All regressions include the stratifying variables (gender, admission test, cohort dummies), and a set of additional controls (own ability, region-of-origin dummies, mother tongue, age, group size). P-values are computed using randomization inference (1,000 draws from the randomization distribution). The table also reports the difference between the coefficients on high- and low-ability peers, as well as the p-value of the difference.
*p $<$ 0.10, **p $<$ 0.05, ***p $<$ 0.01.

Table A.5: Non-linear peer effects on First-year GPA: Effect heterogeneity by own ability

	(1)	(2)	(3)
	Dependent variable: 1st year GPA		
	Own ability high	Own ability middle	Own ability low
Fraction high ability peers	-0.043	0.104	-0.467***
p-value	(0.510)	(0.282)	(0.007)
Fraction low ability peers	0.057	0.103	0.207
p-value	(0.331)	(0.245)	(0.177)
Difference fraction high - low	-0.099	0.002	-0.673***
p-value	(0.324)	(0.526)	(0.000)
N		7,953	

Note: The table presents average marginal effects from an OLS regression of first-year GPA on the fraction of high- and low-ability peers, interacted with a student's own ability (categorized as high, middle, and low). Middle-ability peers are in the reference category. The regression includes the stratifying variables (gender, admission test, cohort dummies), and a set of additional controls (own ability, region-of-origin dummies, mother tongue, age, group size). P-values are computed using randomization inference (1,000 draws from the randomization distribution). The table also reports the difference between the coefficients on high- and low-ability peers, as well as the p-value of the difference.

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

Table A.6: Non-linear peer effects: Effect heterogeneity by own ability

	(1)	(2)	(3)	(1)	(2)	(3)
		Own ability			Own ability	
	high	middle	low	high	middle	low
	Panel A. 1st semester completed			Panel B. 2nd year started		
Frac. high ability peers	-0.015	0.150***	-0.049*	0.007	0.049	-0.093*
p-value	(0.475)	(0.017)	(0.098)	(0.452)	(0.209)	(0.058)
Frac. low ability peers	-0.014	0.126**	0.037	-0.047	-0.018	0.072
p-value	(0.425)	(0.020)	(0.178)	(0.274)	(0.480)	(0.154)
Diff. frac. high - low	-0.001	0.024**	-0.086***	0.054	0.066	-0.165***
p-value	(0.547)	(0.389)	(0.012)	(0.244)	(0.209)	(0.006)
N		8,073			8,073	
	Panel B. Double major			Panel D. Final GPA		
Frac. high ability peers	0.010	0.080	-0.188***	-0.027	-0.145	-0.297
p-value	(0.340)	(0.227)	(0.001)	(0.490)	(0.310)	(0.133)
Frac. low ability peers	0.046**	0.049	-0.118*	0.203	0.267	-0.157
p-value	(0.040)	(0.355)	(0.051)	(0.164)	(0.171)	(0.264)
Diff. frac. high - low	-0.036	0.031	-0.070	-0.229	-0.413*	-0.140
p-value	(0.118)	(0.355)	(0.138)	(0.162)	(0.081)	(0.289)
N		6,343			3,241	

Note: The table presents average marginal effects from regressions of student outcomes on the fraction of high- and low-ability peers, interacted with a student's own ability (categorized as high, middle, and low). Middle-ability peers are in the reference category. The regressions are based on linear models for continuous outcomes and on logit models for binary outcomes. The regressions include the stratifying variables (gender, admission test, cohort dummies), and a set of additional controls (own ability, region-of-origin dummies, mother tongue, age, group size). P-values are computed using randomization inference (1,000 draws from the randomization distribution). The table also reports the difference between the coefficients on high- and low-ability peers, as well as the p-value of the difference.
*p $<$ 0.10, **p $<$ 0.05, ***p $<$ 0.01.

Table A.7: Robustness: Linear-in-means model, without additional controls

	(1)	(2)	(3)	(4)	(5)
	Panel A. First-year grades				
	GPA	Math	Economics	Business	Legal Studies
Peer ability	-0.078**	-0.090**	-0.071*	-0.094**	-0.072*
p-value	(0.039)	(0.021)	(0.051)	(0.019)	(0.053)
N	7,953	7,405	7,848	7,909	7,703
	Panel B. First-year retention				
	1st semester		2nd semester		2nd year
	completed	passed	completed	passed	started
Peer ability	-0.022*	-0.029*	-0.015	0.001	-0.010
p-value	(0.053)	(0.063)	(0.221)	(0.497)	(0.287)
N	8,073	8,073	8,073	8,073	8,073
	Panel C. Major choice				
	Business	Economics	International Affairs	Law	Double major
Peer ability	-0.057**	0.016	0.009	0.002	-0.022**
p-value	(0.022)	(0.255)	(0.326)	(0.468)	(0.018)
N	6,343	6,343	6,343	6,343	6,343
	Panel D. Graduation				
	3 years (on time)	4 years	Final grade		
Peer ability	0.006	-0.024	-0.157**		
p-value	(0.406)	(0.143)	(0.012)		
N	3,953	3,953	3,241		

Note: The table presents average marginal effects from regressions of student outcomes on the average ability in the peer group, using linear models for continuous outcomes, and logit models for binary outcomes. All regressions include the stratifying variables (gender, admission test, cohort dummies), and a reduced set of controls (own ability, group size). P-values are computed using randomization inference (1,000 draws from the randomization distribution).
*p $<$ 0.10, **p $<$ 0.05, ***p $<$ 0.01.

Table A.8: Robustness: Restricted sample

	(1)	(2)	(3)	(4)	(5)
	Panel A. First-year grades				
	GPA	Math	Economics	Business	Legal Studies
Peer ability	-0.088*	-0.036	-0.068*	-0.103**	-0.058
p-value	(0.063)	(0.259)	(0.109)	(0.033)	(0.151)
N	4,992	4,635	4,932	4,968	4,868
	Panel B. First-year retention				
	1st semester		2nd semester		2nd year
	completed	passed	completed	passed	started
Peer ability	-0.013	-0.030*	-0.024	0.001	0.000
p-value	(0.227)	(0.098)	(0.162)	(0.513)	(0.497)
N	5,027	5,027	5,027	5,027	5,027
	Panel C. Major choice				
	Business	Economics	International Affairs	Law	Double major
Peer ability	-0.079***	0.026	-0.007	0.020	-0.028**
p-value	(0.009)	(0.157)	(0.364)	(0.175)	(0.017)
N	3,953	3,953	3,953	3,953	3,953

Note: The table presents average marginal effects from regressions of student outcomes on the average ability in the peer group, using linear models for continuous outcomes, and logit models for binary outcomes. The sample is restricted to cohorts 2003-2009. All regressions include the stratifying variables (gender, admission test, cohort dummies), and a set of additional controls (own ability, region-of-origin dummies, mother tongue, age, group size). P-values are computed using randomization inference (1,000 draws from the randomization distribution).
*p < 0.10, **p < 0.05, ***p < 0.01.

Table A.9: Robustness: Heterogeneity with ability groupings by quartiles of the ability distribution

	(1) 1st year GPA	(2) 1st semester completed	(3) 2nd year started	(5) Double major	(6) Final GPA
Peer ability × high	-0.037	-0.020	0.050	-0.021**	-0.150*
	(0.283)	(0.319)	(0.149)	(0.036)	(0.089)
Peer ability × middle	-0.023	-0.012	-0.011	-0.030	-0.203**
	(0.337)	(0.270)	(0.330)	(0.125)	(0.023)
Peer ability × low	-0.246**	-0.033**	-0.039	-0.011	-0.036
	(0.018)	(0.044)	(0.116)	(0.391)	(0.415)
N	7,953	8,073	8,073	6,343	3,241

Note: The table presents average marginal effects from regressions of student outcomes on the average ability in the peer group, interacted with own ability. Own ability is classified as "high" if the student ranks in the top quartile of the ability distribution, as "low" if the student ranks in the bottom quartile of the ability distribution, and as "middle" otherwise. The regressions use linear models for continuous outcomes, and logit models for binary outcomes. All regressions include the stratifying variables (gender, cohort dummies, admission test), and a set of additional controls (own ability, region-of-origin dummies, mother tongue, age). P-values are computed using randomization inference (1,000 draws from the randomization distribution).
*p $<$ 0.10, **p $<$ 0.05, ***p $<$ 0.01.

Table A.10: Robustness: Non-linearities with ability grouping by ability quartiles

	(1) 1st year GPA	(2) 1st semester completed	(3) 2nd year started	(4) Double major	(5) Final GPA
Fraction high ability peers	-0.105	-0.016	-0.033	-0.051	-0.306**
p-value	(0.168)	(0.333)	(0.230)	(0.076)	(0.040)
Fraction low ability peers	0.072	0.028	0.002	0.014	0.167
p-value	(0.214)	(0.158)	(0.469)	(0.264)	(0.114)
Difference fraction high - low	-0.178*	-0.043	-0.035	-0.065**	-0.472***
p-value	(0.083)	(0.107)	(0.245)	(0.048)	(0.006)
N	7,953	8,073	8,073	6,343	3,241

Note: The table presents average marginal effects from regressions of student outcomes on the fraction of high and low ability peers, using linear models for continuous outcomes, and logit models for binary outcomes. Own ability is classified as "high" if the student ranks in the top quartile of the ability distribution, as "low" if the student ranks in the bottom quartile of the ability distribution, and as "middle" otherwise. Middle-ability peers are in the reference category. All regressions include the stratifying variables (gender, admission test, cohort dummies), and a set of additional controls (own ability, region-of-origin dummies, mother tongue, age, group size). P-values are computed using randomization inference (1,000 draws from the randomization distribution). The table also reports the difference between the coefficients on high- and low-ability peers, as well as the p-value of the difference.
*p < 0.10, **p < 0.05, ***p < 0.01.

A.1 Technical appendix

Randomization inference

Using randomization inference, I test the sharp null hypothesis that peer ability has no influence on the outcome under study. Let θ be the parameter of interest, such that $H_0 : \hat{\theta} = 0$. The realized assignment (or "status quo" assignment) is just one potential assignment out of a large set of counterfactual assignments (or "placebo" assignments), which could have been achieved under the same stratified assignment protocol. Let $\hat{\theta}_{sq}$ denote the effect size under the status quo assignment. The p-value of the randomization test reports the probability of detecting an effect that is at least as large as $\hat{\theta}_{sq}$ in absolute terms just by chance. To determine the p-value, I proceed as follows.

First, I generate a "randomization distribution", which is the distribution of $\hat{\theta}$ under all potentially possible counterfactual assignments coming from the same assignment protocol as the status quo assignment. In order to approximate the randomization distribution, I generate a random subset of all possible counterfactual assignments, keeping the strata proportions in each group fixed. I then generate the treatment variables (e.g., average peer ability) for each of the counterfactual assignments ("placebo treatments") and their corresponding placebo treatment effects $\hat{\theta}$. This results in a randomization distribution of $\hat{\theta}$.

Second, I calculate the p-value of the randomization test of H_0. The p-value indicates the probability of finding an average marginal effect that is in absolute terms as least as large as $\hat{\theta}_{sq}$, if one were to draw a random $\hat{\theta}$ from the randomization distribution.