I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11024

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

Petra Thiemann

SEPTEMBER 2017

DISCUSSION PAPER SERIES

IZA DP No. 11024

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

Petra Thiemann
Lund University and IZA

SEPTEMBER 2017

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany
Phone: +49-228-3894-0
Email: publications@iza.org
www.iza.org

ABSTRACT

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

This paper demonstrates that short-term peer exposure can generate achievement effects which persist for several months and years. I study a mandatory freshmen week for first-year undergraduates and exploit the random assignment of students to freshmen teams. I find that the freshmen week contributes to the formation of persistent social ties. Furthermore, peers' observable characteristics impact college achievement for up to three years. Ability peer effects are non-linear, i.e. very high or low levels of average peer ability in a group harm students' grades. These effects are most pronounced for low-ability students.

JEL Classification: I21, I23, J24

Keywords: peer effects, higher education, natural experiment, gender, region of origin, ability

Corresponding author:
Petra Thiemann
Lund University
Department of Economics
Tycho Brahes väg 1
SE-223 63 Lund
Sweden
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, Scott Carrell, Christina Felfe, Hans Fricke, Guido Imbens, Lance Lochner, Elena Manresa, Teresa Molina, Michel Serafinelli, 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 also was a Postdoctoral Scholar Research Associate at USC Dornsife INET 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. The usual disclaimer applies.

1 Introduction

> *"I enjoyed the freshmen week. [...] You meet new people and then realize, hey, I am not the only one who is anxious. It was very reassuring."* (Participant of a freshmen week.)
>
> *"My freshmen week was terrible, I had a terrible group. [...] I was so alone and asked myself, what am I doing here."* (Another participant of the same freshmen week.)[1]

When students enter college, they encounter many decisions that will influence their future careers: which major to choose, how to form study habits, and whether to stay or to drop out, for example. Peer or reference groups influence these decisions and therefore students' outcomes in important ways (see Epple and Romano, 2011, and Sacerdote, 2011, for reviews).

At the same time, students' networks are particularly malleable during the first weeks at college (c.f. Back et al., 2008). The transition from high school to college disrupts students' old friendship networks and creates new ones. Many students move to a new city, state, or country and thus leave their familiar social environments. On the one hand, students perceive building new networks as particularly challenging at the onset of college education (e.g. Gall et al., 2000). On the other hand, students need friendships in order to integrate well into the college community and mitigate the stress from the life changes that happen during the first few months or weeks at college (e.g. Tinto, 1975).

Aware of the importance of initial interactions, colleges around the world offer short interventions that provide students with opportunities to form new friendships and social connections. These interventions are often "orientation weeks" (or "freshmen weeks"), during which the students explore campus life in groups. While longer-term peer exposure has been intensely studied (c.f. Sacerdote, 2001, Lyle, 2007, 2009, Carrell et al., 2009, 2013, Oosterbeek and van Ewijk, 2014, Booij et al., 2017, Feld and Zölitz, 2017), it is unclear whether short-term peer exposure generates lasting effects on college achievement. Furthermore, little is known about how to design short-term interventions and how to compose peer groups in order to positively affect students' outcomes.

[1] Quotations from interviews of undergraduate students at the University of St. Gallen carried out in the course of a study on the reasons for dropout (authors: Stephan Egger and Diana Reiners).

Such short-term peer interventions, however, can potentially provide a cost-effective and easily implementable way to integrate students into the college community.

This paper provides new causal evidence on the persistent effects of short-term peer exposure in a college setting on social tie formation and student performance. I exploit the exogenous assignment of students to orientation week groups to study two questions: (1) Does the peer group assignment persistently impact friendship formation? (2) Do the characteristics of group members play a role for students' grade outcomes and retention? In particular, detailed information on the place where the students completed their high school degree allows me to explore the composition of the peer groups in terms of students' region of origin.

As students typically self-select their friends and study partners, it is in most contexts difficult to determine whether similar outcomes among peers are a result of peer influence or self-selection. To address this problem, this paper exploits the random assignment of students to small orientation week groups at a Swiss university. The students spend a 60-hour program in groups of 15 students. In the course of an incentivized case study competition, which takes place during the week, the group members get to know each other through group work. The randomized group assignment allows me to disentangle the causal effect of first-week encounters from other student characteristics that impact student performance and friendship formation.

I use a comprehensive administrative data set, which contains a range of student background characteristics and detailed information on academic achievement (grades in mandatory first-year courses, enrollment, dropout, and graduation records) for 8,000 undergraduates (cohorts 2003-2012). 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 from the freshmen week, and data on student participation in mandatory review sections.

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

In the second part of the paper I show that while gender peer effects are moderate, peers' region of origin and ability impact performance even in the longer run. Students who completed their high school degree abroad (here: outside Switzerland) have a significantly negative impact on the performance of their peers. If the share of peers from abroad in a peer group increases by 10 percentage points, a student's first-year GPA decreases by 0.04 standard deviations on average. To explore the effect of peer ability, I use an imputed measure of peer ability, which is a weighted combination of a set of characteristics that predict ability. I find an inverse U-shaped relation between peer ability and performance, i.e. being in a group with a very high or with a very low average ability level can harm a student's performance. For example, for students in the bottom 10% of the peer ability distribution, increasing peer ability by one standard deviation leads to an increase in first-year GPA by 0.07 standard deviations, a non-negligible effect. Low-ability students are most vulnerable to being in groups with extremely high or extremely low peer ability, indicating that they benefit from positive peer spillovers, but might also suffer from discouragement effects in the presence of high-ability peers.

Furthermore, I show results on the link between ability and region-of-origin peer effects. Since students who completed their high-school degree abroad are high-performing students on average, the effect of ability and region of origin are intertwined. I thus explore these characteristics both separately and simultaneously. I find that depending on the outcome at hand, either characteristic can gain importance. While peers' ability matters relatively more for GPA, an outcome that strictly measures performance, peers' region of origin proves relatively more important for first-semester completion, an outcome that is to a larger extent choice-based.

This paper contributes to the literature in two main respects. First, the paper shows that short-term peer interventions can be an effective way to generate longer-term social ties. Prior studies have focused on longer-term interventions, which typically last for one semester or year. These interventions include, for example, assignments to work groups, study groups, dorm rooms, and military squadrons (Sacerdote, 2001, Lyle, 2007, 2009, Carrell et al., 2009, 2013, Oosterbeek and van Ewijk, 2014, Booij et al., 2017, Feld and Zölitz, 2017). Assessing a short-term intervention thus proves important since it can be readily organized and manipulated in many educational institutions, without interfering with standard curricular, course choices, or living arrangements. In fact, orientation weeks are already implemented in many universities and colleges around the world. Moreover, using a cost-effective short-term intervention to help individuals integrate into a new community can prove important in a variety of settings, a topic that deserves further study.

Second, this paper demonstrates the link between multiple peer characteristics, and that studying this link is important for understanding the channels that drive peer effects. The present study relates both to the literature on region-of-origin peer effects at lower levels of schooling (see, e.g., Ohinata and Van Ours (2013) for a review) and to the literature on ability peer effects (see, e.g., Carrell et al., 2009, 2013 and Booij et al., 2017 for ability peer effects at the college level). Since these characteristics confound each other to some extent, conclusions on the impact of peer ability and region of origin cannot easily be derived from studying each of the characteristics in isolation. Studying multiple peer characteristics jointly, however, bears methodological complications that have yet to be resolved (c.f. Graham, 2011). This paper presents different ways to do so, based on a standard approach (including a set of peer characteristics as explanatory variables into a regression), a novel approach using an imputed ability measure that summarizes several peer characteristics into a score, and a combination of the two approaches. Showing that region-of-origin peer effects can be partly explained by ability peer effects, but also impact outcomes in their own right, this study points out that investigating multiple peer dimensions enriches the analysis and ultimately can lead to more refined conclusions.

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 and discussed in light of the literature in Section 6. Section 7 concludes.

2 Setting

2.1 The Freshmen 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 minor exceptions.[2] Academic performance by the end of the first year determines whether 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 final exams.[3] After the first year, students choose their major.

All undergraduate degrees start with a mandatory freshmen week. This week familiarizes the students with the university's infrastructure (e.g. library and online tools) and facilitates contacts between the students. The students are sorted into teams at the beginning of the week. Each team consists of 16 students on average, with between 56 and 60 teams per cohort. Group sizes and the number of groups can vary between cohorts. Moreover, variations of group sizes within cohorts emerge from organizational constraints (i.e., room size). During the week, students spend approximately 60 hours in their groups, with about 75% of the time dedicated explicitly to team activities (see Figure A.1 in the appendix). A case study competition between the groups forms the core team activity.

The assignment mechanism to freshmen groups ensures random assignment of peers in terms of observable and unobservable characteristics, conditional on gender and the admission requirement that applies to the respective student: all individuals with a non-Swiss high school diploma and a non-Swiss nationality have to complete an admission exam; all other students are unconditionally admitted.

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

[3] Students who fail can repeat all first year courses in order to be admitted to the second year.

To randomize students into groups, the university administration 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 ranked 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 4 strata. Afterwards, an administrator re-distributes some students across groups in order to match the group sizes to available room sizes. This redistribution occurs unsystematically. Section 4.3 shows that the resulting groups are balanced in terms of observable student characteristics.

2.2 Persistent impact on friendships and study partnerships

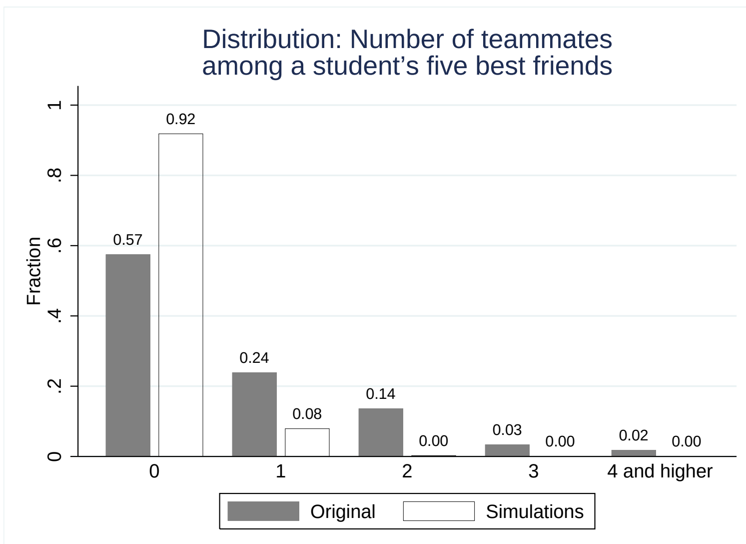
I use two data sources to demonstrate that the freshmen week persistently impacts the formation of friendships and study partnerships among undergraduate students. First, I conduct an online survey of a cross section of undergraduate students, who were in their second to fourth year at the time of the survey (May 2012).[4] This survey focused on the determinants of friendships and study partnerships of undergraduate students. The survey contained a question on how many of a students' five best friends or study partners were teammates from his or her freshmen week.[5] Second, I study how students select themselves into economics review sections. Students can choose one out of 36 sections. All sections are identical in content and all take place on the same weekday, but they differ with respect to the daytime and the instructor. The average number of participants in a review section is 14, with a minimum of six and a maximum of 21 students. I use attendance data from the first review section, which takes place one week after the end of the freshmen week, to detect clustering of teammates in the same review sections.

As Figure 1 depicts, teammates are over-represented among a student's five best friends even several months or years after the freshmen week. In the sample of 380 survey respondents, the av-

[4]The online survey was administered to all undergraduate students of cohorts 2008-2010 who were still enrolled of the time of the survey (2,124 students) and had a response rate of 18%.

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

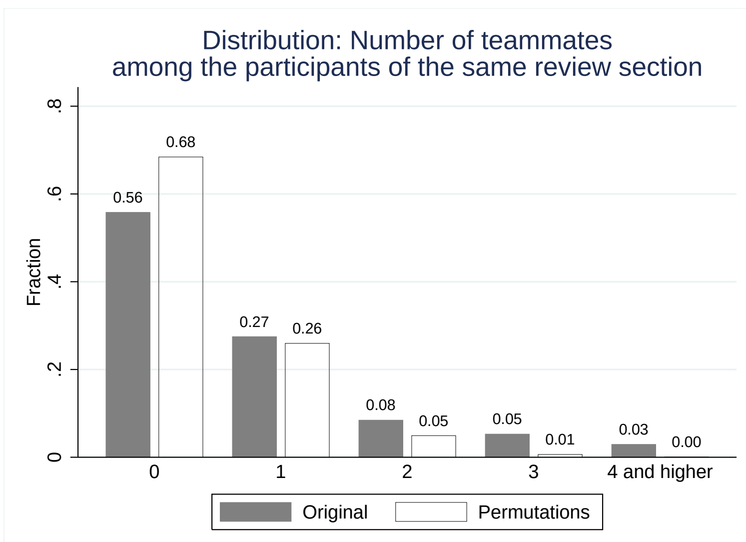
Figure 1: Friendship formation



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 ($p < 0.001$). Source: Own calculations based on survey data and administrative data from the University of St. Gallen.

erage 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. 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. I then average across all 1,000 allocations. 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 ($p < 0.001$). These results are, however, based on a non-representative survey, which includes roughly 13% of students.[6] Thus, I perform a similar analysis based on how students select into review sections.

Figure 2: Selection into the same review sections



Note: The dark bars show the actual distribution of the number of teammates from the freshmen week in a student's review section, based on participation data. The white bars show the same distribution, based on 1,000 simulations under the assumption that students participate in review sections independent of other members of their freshmen team. The data contain all review-section 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 review section is 0.78 in the original data and 0.39 on average across 1,000 permutations. The difference is significant ($p < 0.001$). Source: Own calculations based on administrative and participation data from the University of St. Gallen.

[6]Ideally, both distributions would be to based on the same sample of students. I cannot, however, merge the administrative data with the survey data to filter out survey non-participants from the administrative data.

Figure 2, which is based on student participation in review sections, reveals a similar picture: teammates are over-represented among members of the same review section. To summarize this, I depict the distribution of teammates (excluding the student himself) among the participants of his review section. I also create 1,000 counterfactual allocations of the students across review sections, holding the size of the review sections fixed. I then compare the original distribution to the simulated distribution. The average number of teammates in a student's review section 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-week friendships. Freshmen teammates do not only bond during the first week, but also deepen their bond as they participate in the same activities, such as academic review sections. At the same time, not all teammates matter. The majority of students 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.[7]

3 Data and descriptive statistics

3.1 Sample

The dataset consists of administrative records for 8,073 freshmen who started their undergraduate degree between 2003-2012.[8] Background (pre-treatment) characteristics as well as outcomes are computed from enrolment 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

[7]Unfortunately, the student populations from the two data sources (survey and review section data) does not overlap. Thus, I cannot study whether the students who have persistent friendships with teammates are also the students who tend to select into the same review sections with their teammates.

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

groups.[9] 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: age, gender, whether students had to complete an admission test, mother tongue (German vs. non-German), nationality (Swiss, German/Austrian, other), 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).

Table 1 presents descriptive statistics of the student body. First, two thirds of the students are male. Second, most students are either Swiss (76%) or have a Swiss high school degree (77%). Only 18% of students neither are Swiss nor have a Swiss high school degree, and therefore needed to pass an admission test. While all other students are unconditionally admitted, this test allows the university to meet the admission quota for foreign students, which is legally restricted to at most 20%.[10] Because of this selective process, the admission test takers consistently outperform all other undergraduate students. Third, although Switzerland has four different official languages (German, French, Italian, Romansh), only 11% of the students speak a non-German mother tongue, presumably because German is the main language of undergraduate instruction for the cohorts under study.

To capture the diversity of the student body in terms of the region of origin of a student, I define six categories. The first four categories contain all students who completed their high school diploma in Switzerland: The students in the first category are from the region of "East Switzerland", which is the region surrounding and to the east of St. Gallen.[11] Most of these students grew up in commuting distance from the University and form the second largest group of first-year un-

[9]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.

[10]This is a decision of the canton (state) of St. Gallen, who provides the main source of funding to the university.

[11]The corresponding cantons are: Appenzell-Innerrhoden, Appenzell-Ausserrohden, Glarus, St. Gallen, Schaffhausen, Thurgau.

Table 1: Descriptive statistics of the estimation sample

	Mean	SD	Min.	Max.	Obs.
Student demographics					
Female	0.32	0.47	0	1	8,073
Age	20.14	1.87	16	48	8,073
Non-Swiss nationality ("foreign")	0.24	0.43	0	1	8,073
Non-German mother tongue	0.11	0.31	0	1	8,073
Region of origin (location of high school degree)					
Region 1: Close to the university (German-speaking)	0.25	0.43	0	1	8,073
Region 2: Remaining German-speaking cantons	0.36	0.48	0	1	8,073
Region 3: Mixed-language cantons	0.11	0.31	0	1	8,073
Region 4: Non-German speaking cantons	0.06	0.24	0	1	8,073
Region 5: Swiss, HS degree foreign	0.05	0.21	0	1	8,073
Foreign, HS degree foreign	0.18	0.38	0	1	8,073
Group variables					
Group size	16	3	7	22	8,073
Fraction: Female	0.32	0.07	0	0.57	8,073
Fraction from region 1	0.25	0.11	0	0.71	8,073
Fraction from region 2	0.36	0.13	0	0.80	8,073
Fraction from region 3	0.11	0.08	0	0.42	8,073
Fraction from region 4	0.06	0.06	0	0.36	8,073
Fraction from region 5	0.05	0.05	0	0.3	8,073
Fraction foreign, HS degree foreign	0.18	0.07	0	0.5	8,073
Outcomes					
First semester: all courses completed	0.91	0.28	0	1	8,073
First semester: passed	0.79	0.41	0	1	8,073
Second semester: all courses completed	0.74	0.44	0	1	8,073
Second semester: passed	0.66	0.47	0	1	8,073
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
Completed degree within 3 years (on time)	0.22	0.42	0	1	5,027
Completed degree within 4 years	0.64	0.48	0	1	5,027

Note: The table presents descriptive statistics of the estimation sample (cohorts 2003-2012), based on administrative records. First-year GPA 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 are only reported for those students with a valid record in the respective course. Grades are standardized by cohort. Data on degree completion is only available for cohorts 2003-2009 because of censoring. Source: Own calculations based on administrative data from the University of St. Gallen.

dergraduates (25%). These students are thus able to stay closely connected to their networks at home, and the University of St. Gallen may serve as their most convenient choice. Moreover, all of these students spoke German at their high school. The students in the second category come from all other German-speaking cantons.[12] With 36% of all undergraduates, this category represents the largest regional group. The students in the third category are from cantons with at least two official languages,[13] and account for 11% of the student body. Some of these students grew up bilingually, or were exposed to two languages in their daily lives; but German may not be their mother tongue. The students in the fourth category originate from cantons in which either French or Italian is the language of instruction in high school.[14] These students on average experience a language barrier to studying in St. Gallen. Because of this barrier, these students are a minority among the first-year undergraduates (6%).

Two further groups consist of students who completed their high school degree outside Switzerland: First, 5% of the students are Swiss citizens with a high school degree from abroad; 80% of these students have a German mother tongue. Second, 18% of undergraduates are non-Swiss citizens who also obtained their high school degree abroad and thus had to complete an admission test. 88% of these students are either German or Austrian, and 93% of these students have a German mother tongue.

3.3 Peer characteristics

As group composition or peer variables I use the fraction of female teammates and the fraction of teammates from each of the regions of origin (proxied by the region in which a student obtained his/her high school degree). Each of these variables are computed as leave-own-out means, i.e., excluding the student him-/herself. The fractions in the minority categories (the fraction of students from non-German-speaking cantons and the fraction of Swiss students who obtained their high school degree abroad) are with 6% and 5% negligibly small (see Table 1). Thus, the treatment

[12]Cantons: Aargau, Baselland, Basel city, Lucerne, Nidwalden, Obwalden, Solothurn, Schwyz, Uri, Zug, Zurich.

[13]Cantons: Berne, Fribourg, Vallais (official languages: German and French), Grisons (official languages: German, Italian, and Romansh).

[14]Cantons: Geneva, Jura, Neuchâtel, Ticino, Vaud.

variation in these variables is limited. In all regressions, I use the largest category, students from region 2, as the reference category.

3.4 Outcomes

The set of outcome variables contains information on students' performance for up to four years after the students started their undergraduate degree.

A set of binary variables captures persistence, retention, and dropout behavior: whether students complete all compulsory courses during the first and second semester of the first year, whether they pass the first and second semester of the first year in their first attempt (all others have to repeat the full first year or to drop out), and whether they graduate within three years (on time). As the first year is competitive, only two thirds of students manage to pass the first year in the first attempt. An almost equal fraction of students, however, manages to graduate in at most four years—which indicates that the first year is the most difficult barrier to graduation.

A further set of variables represents students' grades. 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 selected first-year courses. To minimize missing outcomes, I choose four mandatory first-semester courses that were completed by most students. All grades are standardized at the cohort level.

The performance and persistence variables are highly correlated (see also Tafreschi and Thiemann, 2016, who use the same administrative data from St. Gallen). While I use the full set of outcomes in the main part of the analysis (Section 5.1), I focus on GPA in the sections on robustness, effect heterogeneity, and channels (Sections 5.2-5.4).

4 Empirical approach

4.1 Empirical model

Following Manski (1993), I implement a linear-in-means model of peer effects. Consider a student's GPA or course grade, which depends on both individual and group characteristics:

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

where Y_{igc} is the GPA of student i in cohort c who is a member of freshmen group g, X_i is a vector of pre-determined characteristics of student i, and the vector $\overline{X}_{-ig}$ represents the average of these characteristics across 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 indicators, W_i is a vector of additional individual-level controls that are not in X_i, and ϵ_{igc} is an idiosyncratic error term.

If the groups are randomly assigned, the γ-coefficient vector 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 added covariates (groupsize, cohort dummies, further individual controls) fulfil two main purposes. First, the covariates account for the stratification of the randomization with respect to gender and the admission requirement. While an indicator variable for being female is included in the X_i-vector, an admission test indicator is included in the W_i-vector. Second, the covariates capture changes in group sizes and cohort composition over time that may be correlated with changes in student outcomes. Thus, the regression includes both group sizes and cohort dummies.

The conditional expectation of test scores takes the form:

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

In addition to continuous outcomes (GPA and course grades), I consider a set of binary outcomes (e.g. whether the student graduates on time). For these outcomes, the conditional probability takes the form:

$$P[Y_{igc}|X_i, \overline{X}_{-ig}, s_g, D_c, W_i] = \alpha + X_i'\beta + \overline{X}'_{-ig}\gamma + \xi s_g + D_c'\delta + W_i'\rho. \tag{3}$$

The parameter of interest is the average marginal effect of increasing the fraction of students of a certain subgroup (e.g. the fraction of female students). Let $\overline{x}_{-igk}$ denote the fraction of students with characteristic k in student i's group. I define the average marginal effect as

$$\theta_k = E\left[\frac{\partial \mathrm{E}[Y_{igc}|X_i, \overline{X}_{-ig}, s_g, D_c, W_i]}{\partial \overline{x}_{-igk}}\right] \tag{4}$$

for the continuous dependent variables, and as

$$\theta_{k'} = E\left[\frac{\partial \mathrm{P}[Y_{igc}|X_i, \overline{X}_{-ig}, s_g, D_c, W_i]}{\partial \overline{x}_{-igk'}}\right] \tag{5}$$

for the binary dependent variables.

The policy interpretation of the parameters of interest depends on the definition of peer variables and reference categories. In the case of gender, a binary variable, the effect is the increase in the fraction of females, which corresponds to an equal reduction in the fraction of males. In the case of the region of origin, which consists of multiple categories, all average partial effects have to be interpreted relative to the reference category (here, the proportion of students from region 2, which is the German-speaking part of Switzerland, except for the cantons close to St. Gallen). For example, the average marginal effect of an increase in the fraction of students from outside Switzerland corresponds to a policy experiment in which the fraction of students from region 2 is reduced, and the fractions of students from all other regions remain unchanged.

4.2 Estimation and inference

To compute the parameters of interest, I proceed in two steps. First, I estimate the coefficients from Models (2) and (3). To estimate Model (2) I use ordinary least squares; to estimate Model (3), I assume a probit specification and carry out a maximum likelihood estimation. Then, by plugging in the estimated model parameters, I compute the sample analogues to Equations (4) and (5).

The p-values are based on randomization inference. This inference method exploits the randomization mechanism directly. For each of the peer characteristics, I test the sharp null hypothesis that this characteristic has no influence on the outcome, i.e. $H_0 : \hat{\theta}_k = 0$. The realized assignment (or "status quo") is just one potential assignment out of a large set of counterfactual assignments (or "placebos") , which could have been achieved under the same stratified randomization scheme. Denote the effect size under the status quo as $\hat{\theta}_{k_{sq}}$. The p-value of the randomization test reports the probability of detecting an effect that is at least as large as $\hat{\theta}_{k_{sq}}$ in absolute terms just by chance. To determine the p-value, I proceed in three steps. First, I define the "randomization distribution" as the distribution of $\hat{\theta}_k$ under all potentially possible counterfactual assignments coming from the same randomization scheme as the status quo assignment. Second, 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 compute $\hat{\theta}_k$ for each of the counterfactual assignments, using Models (2) and (3). Third, 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 $\hat{\theta}_{k_{sq}}$, if one were to draw a random $\hat{\theta}_k$ from the randomization distribution.

For comparison, I also report analytic standard errors based on the the delta-method in Section 5.2. As both inference methods result in similar levels of confidence, I proceed by presenting analytic standard errors in Section 5.4.

4.3 Test for random assignment

According to the administrative rules, the freshmen groups are randomly assigned, conditional on a student's gender and admission requirement. Thus, individual student characteristics should be unrelated to the composition of students' peer groups, conditional on these variables. To test this assumption, I regress the treatment variables on student background characteristics. The specifications include the same set of control variables as used in Models (2) and (3). Table A.1 in the appendix presents the coefficients from these regressions.

Out of the 49 tested parameters, only two are significant at the 10%-level, and only one is significant at the 1%-level. This is a likely result under the assumption of random assignment.[15]

[15]If the significance levels of the coefficients were independent of each other, the probability of at least one out of 49 coefficients being significant at the 1%-level and at least two out of 49 coefficients being significant at the 10%-level is 0.39, a number that is above any conventional significance level (based own simulations with 100,000 random draws).

Thus, the test does not reject the random assignment assumption.

5 Results

5.1 Main results

Tables 2 and 3 present the effect of peer composition in terms of gender and region of origin on academic achievement. Table 2 displays the results for semester completion and retention, and Table 3 shows the results for first-year GPA as well as individual course grades.

Gender peer effects in both achievement and retention during the first year are small and insignificant, but matter more toward the end of a student's degree. While a higher share of females in a group has a slightly negative effect in the first semester, the sign turns positive in the second semester, but in both cases the effects are insignificant. Yet, a higher share of female peers negatively and significantly affects the probability to graduate on time. An increase in the share of females by 10 percentage points leads to a decrease in the probability of graduating on time by about 2 percentage points (10 percent with respect to the baseline of 22 percent). This effect is large and may be driven by the low performance of females, compared to males. On average, females obtain grades that are 0.11 standard deviations lower than the grades of males with comparable characteristics (see Table A.2).[16]

While the gender composition of the freshmen teams matters only during later stages of a student's undergraduate studies, the reverse holds for the team composition in terms of the region of origin of the peers. The region of origin of a student's peers matters especially during the first semester of a students career, yet fades out toward the second semester and is no longer significantly different from zero at that point. The group of peers that accounts for most of the effect are students who obtained their high school degree outside Switzerland (both Swiss and non-Swiss citizens). A higher fraction of peers with foreign high school degrees leads to lower rates of course

[16]It is unclear why men outperform females in the given setting. This result might for example be driven in differences in unobservable characteristics between males or females (e.g., differences in skills, preferences, or perceptions, and may be further exacerbated by the environment (e.g. through the competitiveness of the environment).

completion during the first and second semester, and lower rates of passing the first semester, which is again partly due to course non-completion. Moreover, peers with foreign high school degrees affect grades during the first semester and lead to a lower first-year GPA. Increasing the share of students who obtained their high school degree outside Switzerland by 10 percent decreases GPA by between 2.4 and 3.7 percent of a standard deviation.

By contrast, students who completed their high school degree from non-German speaking cantons impact student retention and achievement positively. The students from non-German speaking cantons may be positively selected on their motivation to pursue their studies in a foreign language. Given that the majority of courses are held in German, most of these students experience a language barrier, for example when taking exams, writing papers, giving presentations, and communicating with peers and instructors. Motivation spillovers may thus lead to the positive peer effects of this group. Such motivation effects would also explain why this group generates highly significant effects on outcomes that are to a larger extent based on choices (i.e., persistence during the first year), but no significant effects on purely achievement-based outcomes (i.e., course grades).

5.2 Robustness of the main results

The results are robust to alternative inference methods (analytic standard errors instead of randomization inference) as well as different functional forms. Tables A.5 and A.6 display results for regressions of GPA and the probability of passing the first semester on peer characteristics and report analytic standard errors.

The p-values increase slightly under analytic inference (Tables A.5 and A.6), compared to the p-values under randomization inference (Tables 2 and 3). This is intuitive as the randomization inference accounts only for the variation that stems from the (random) group assignment, and does not take sampling variation into account. However, the sample contains 97% of the population of students in St. Gallen in the years under study. Therefore, the analytic standard errors, which account for sampling variation, may be overly conservative.

Moreover, based on F-tests I study whether adding second order polynomials of peer characteristics alters the significance of peer characteristics (Tables A.5 and A.6). The joint significance

Table 2: Main results: First-year outcomes and on-time graduation

	(1)	(2)	(3)	(4)	(5)
	Outcomes: student retention				
	First semester		Second semester		Graduation
	completed	passed	completed	passed	on-time
Peers' gender					
Fraction female					
	-0.015	-0.017	0.010	0.068	-0.193**
	(0.316)	(0.417)	(0.452)	(0.236)	(0.017)
Peers' region of origin					
Fraction region 1					
(close to St. Gallen)	-0.037	-0.026	-0.051	-0.015	-0.012
	(0.108)	(0.292)	(0.142)	(0.387)	(0.401)
Fraction region 3					
(mixed-language canton)	-0.004	-0.042	-0.031	0.020	0.021
	(0.441)	(0.244)	(0.320)	(0.403)	(0.374)
Fraction region 4					
(non-German speaking canton)	0.009	0.179***	0.123*	0.101	-0.032
	(0.411)	(0.009)	(0.057)	(0.120)	(0.348)
Fraction region 5					
(Swiss, HS degree abroad)	-0.061	-0.166**	-0.152**	-0.095	-0.074
	(0.124)	(0.016)	(0.045)	(0.162)	(0.263)
Fraction foreign					
(foreign, HS degree foreign)	-0.089**	-0.093*	-0.089	-0.026	-0.098
	(0.018)	(0.076)	(0.103)	(0.355)	(0.165)
Obs.	8,073	8,073	8,073	8,073	5,027

Note: The table presents average marginal effects from regressions of student outcome variables on group composition variables. The p-values are based on randomization inference (1,000 random draws from the randomization distribution). The regressions are based on probit models and estimated using MLE. All regressions include cohort dummies and controls (gender, region of origin, admission test indicator, age, group size). Source: Own calculations based on administrative data from the University of St. Gallen.
*p < 0.10, **p < 0.05, ***p < 0.01.

Table 3: Main results: GPA and course grades

	(1)	(2)	(3)	(4)	(5)
	Outcomes: First-year grades				
	GPA	Math	Business Admin.	Economics	Legal Studies
Peers' gender					
Fraction female					
	0.034	0.141	-0.054	-0.047	0.106
	(0.452)	(0.164)	(0.354)	(0.335)	(0.319)
Peers' region of origin					
Fraction region 1					
(close to St. Gallen)	-0.115	-0.115	-0.122	0.021	-0.036
	(0.156)	(0.172)	(0.121)	(0.401)	(0.380)
Fraction region 3					
(mixed-language canton)	0.008	-0.152	-0.085	0.011	-0.014
	(0.472)	(0.156)	(0.269)	(0.451)	(0.451)
Fraction region 4					
(non-German speaking canton)	0.153	0.213	0.085	0.210	0.198
	(0.181)	(0.122)	(0.297)	(0.118)	(0.151)
Fraction region 5					
(Swiss, HS degree abroad)	-0.367**	-0.284*	-0.382**	-0.110	-0.406**
	(0.029)	(0.084)	(0.022)	(0.289)	(0.021)
Fraction foreign					
(foreign, HS degree foreign)	-0.244*	-0.302**	-0.215*	-0.156	-0.134
	(0.055)	(0.027)	(0.071)	(0.141)	(0.166)
Obs.	7,953	7,405	7,848	7,909	7,703

Note: The table presents average marginal effects (AMEs) from regressions of student outcome variables on group composition variables. The p-values are based on randomization inference (1,000 random draws from the randomization distribution). The coefficients are estimated using OLS. All regressions include cohort dummies and controls (gender, region of origin, admission test indicator, age, group size). Source: Own calculations based on administrative data from the University of St. Gallen.
*p < 0.10, **p < 0.05, ***p < 0.01.

of the peer characteristics changes only slightly in response to the specification of the functional form. With respect to GPA, the variables that capture peers' region of origin are jointly insignificant, independent of the functional form (Table A.5, columns 2 and 3). With respect to passing the first semester, the variables that captures peers' region of origin are jointly significant, independent of the functional form (Table A.5, columns 2 and 3).

5.3 Effect heterogeneity

The analysis of effects for the whole sample may mask important heterogeneity in the effects, which may then inform reallocation policies such as grouping according to gender or region of origin (Graham et al., 2010, Graham, 2011).

I find no strong indication for effect heterogeneity by gender (Table A.3). The fraction females seems to positively affect females, and to negatively affect males. But none of the effects is significant, and the coefficients are also not significantly different across the two groups. However, the signs of the effects would be in line with a model where students benefit from same-gender interactions and friendships more than from mixed-gender interactions. Moreover, I find little evidence for gender heterogeneity in region-of-origin peer effects. The most pronounced contrast is in the effect of students from non-German-speaking cantons, which is only positive and significant for females.

Table A.4 shows effect heterogeneity by a student's region of origin. Students do not benefit from same-region peers. Moreover, students who completed their high school degree abroad generate a negative impact on all students, with the exception of students from mixed-language cantons (region 3). Students from mixed-language cantons are particular in that they benefit from all peers, except from the peers that are also from mixed-language cantons.

In sum, the effect heterogeneity patterns do not suggest any specific way to distribute students across groups in order to maximize student achievement. In general, with multiple peer characteristics, reallocation policies are difficult to derive based on regression tables alone.

5.4 Ability as a channel

5.4.1 Ability peer effects

One potential channel for the peer effects shown in the previous sections may be ability peer effects. Specifically, non-Swiss students with high school degrees from outside Switzerland are highly selected and thus high-performing students. They outperform most of their peers during the first year: According to the regressions in Table A.2, students who pass the admission test outperform the reference group—students from the German-speaking part of Switzerland—by 0.75 standard deviations on average by the end of the first year. Being exposed to these high achievers could hinder the performance of lower-achieving students, for example because of a discouragement effect (Hoxby and Weingarth, 2005, Carrell et al., 2013, Tincani, 2014, Feld and Zölitz, 2017).

By contrast, Swiss students who obtained their high school degree from abroad are the lowest performing students among the regional groups. These students neither had to complete their high school degree within the selective Swiss high school system,[17] nor did they have to pass an admission test, and are therefore the least selected students. They perform on average 0.38 standard devations below the reference group. These students may generate negative peer effects as well (Carrell et al., 2009, 2013, Tincani, 2014, Feld and Zölitz, 2017).

Being exposed to students with extremely high or extremely low ability levels could thus explain the pattern of regional peer effects. To further investigate this claim, I analyse peer effects based on an imputed measure of peer ability. I start by predicting each student's ability using the coefficients from a leave-own-year-out regression of first-year GPA on student background characteristics, including the student's region of origin and gender. I standardize the imputed ability measure to have a mean of zero and a standard deviation of one in each cohort. I then verify the validity of this ability measure by comparing 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 (Frederick, 2005) and was collected in the course of a classroom experiment (see Schulz et al., 2016, for details of the data collection). The correlation between imputed

[17]For example, in 2007, only 21% of Swiss students graduated from high school with a degree that allowed them to enroll in a university ("Matura"). By contrast, 41% of German students completed a degree that allowed them to enroll in a university (OECD, 2009).

ability and cognitive reflection is 0.27 (p-value < 0.0001, based on 1,627 observations). Finally, I compute the leave-own-out measure of peers' imputed ability for each student.

The results from the estimations of various functional forms of peer ability suggest that ability peer effects follow a quadratic functional form (see Table 4 in the appendix) and are inverse U-shaped. To illustrate this point, Figure A.2 in the appendix represents the average structural function (Panel 1) and its derivative (Panel 2) within the support of peer ability (see, for example, Blundell and Powell, 2004). All else equal, first-year GPA is highest for students in the mid-range of peer ability and lowest for students at the extremes of peer ability. For example, for students in the bottom 10% of the peer ability distribution, increasing peer quality by one standard deviation (or 0.22) corresponds to an increase in first-year GPA by 0.02 standard deviations on average. Similarly, for students in the top 10% of the peer ability distribution, decreasing peer quality by one standard deviation leads to an increase in first-year GPA by 0.07 standard deviations.

Table 4 investigates effect heterogeneity with respect to a student's own ability (Column 4). To this end, students are classified into high-, middle-, and low-ability students, based on their position in the imputed ability distribution (lowest, middle, and highest tercile). The results suggest that peer ability matters most for the low-ability students. First, the coefficients are highly significant only for this group. Second, the effect of peer ability is overall negative, suggesting that these students would benefit from lower levels of peer quality on average. Third, the second-order term is largest for the low-ability students, i.e., groups with extremely high and extremely low ability harm low-ability students most strongly, compared to middle- and high-ability students. The inverse U-shape pattern, however, holds for high-, middle-, and low-ability peers alike.

As the inverse U-shaped pattern suggests, a principal who cares about average outcomes may create an allocation in which students with different ability levels are evenly distributed across groups. The present random allocation should be close to this goal. A back-of-the-envelope calculation suggests that in the aggregate, average GPA can be increased by 0.02 standard deviations if a principal created a completely balanced mix of students in every group;[18] but the implementation of such an allocation may not be entirely feasible because of institutional constraints and small

[18]To perform this back-of-the-envelope calculation, I set each student's peer ability to 0 (i.e. to the mean ability in the population) and compute the fitted value from regression (2) in Table 4.

Table 4: Effects of peer ability (imputed)

Peer ability (imputed)	-0.081	-0.111**	-0.133*	
	(0.050)	(0.050)	(0.079)	
Peer ability (squared)		-0.419***	-0.396**	
		(0.159)	(0.171)	
Peer ability (cubic)			0.135	
			(0.438)	
Low X Peer ability				-0.243***
				(0.092)
Low X Peer ability (squared)				-0.609**
				(0.255)
Middle X Peer ability				0.000
				(0.085)
Middle X Peer ability (squared)				-0.350
				(0.220)
High X Peer ability				-0.087
				(0.093)
High X Peer ability (squared)				-0.308
				(0.288)
Controls	Yes	Yes	Yes	Yes
Cohort effects	Yes	Yes	Yes	Yes
Adjusted R-squared	0.082	0.083	0.083	0.083
Number of observations	7,953	7,953	7,953	7,953

Note: The table presents the regression coefficients from OLS regressions of student first-year GPA on the average ability in the peer group. The ability measure is imputed based on leave-own-year-out regressions of first-year GPA on student characteristics (gender, region of origin, admission test indicator, mother tongue, nationality, age, age squared, cohort dummies). See Section 5.4 for details on the computation of the ability measure. Panel (4) shows effect heterogeneity. Each student is classified as high-, middle-, or low-ability student, according to his/her position in the ability distribution (top, middle, or bottom tercile of the distribution), and the treatment variables are interacted with the respective indicator variables. Analytic standard errors, clustered at the group level, are in parentheses. The regressions include control variables (gender, region of origin, admission test indicator, mother tongue, nationality, age, age squared, imputed ability, cohort dummies, and group size). Source: Own calculations based on data from the University of St. Gallen.
*p < 0.10, **p < 0.05, ***p < 0.01.

group sizes.

The reallocation problem becomes more complicated when effect heterogeneity with respect to a student's own ability is taken into account. In the present setting, no analytic solution for this reallocation problem exists, because of the high dimension of the problem. Further research, which likely will involve simulation methods, is needed to address such allocation problems.

5.4.2 Disentangling ability peer effects from region-of-origin and gender peer effects

In order to disentangle the peer ability effect from the effects of peers' gender and peers' region of origin, Tables A.5 and A.6 display the results of different models. The different models contain either ability peer effects (columns 3-4), region-of-origin and gender peer effects (columns 1-2), or all peer effects (columns 5-6). For example, if students from different regions have different characteristics that are independent of ability (e.g., different motivation levels or cultural backgrounds), then only part of the effect of regional composition would be reflected in the ability peer effect.

Ability is the main driver of peer effects in GPA, and the effect size hardly changes when including additional variables (Table A.5). Peer ability has a significant effect on GPA not only in a model with only peer ability (column 4), but also in models with ability, gender, and region-of-origin peer effects combined (column 6). In particular, including region-of-origin variables changes the significance level of ability peer effects only slightly (p-value of the F-test changes from 0.020 to 0.017 when moving from column 4 to column 6). By contrast, the region-of-origin peer effects are never jointly significant, neither in the model without ability peer effects (columns 1-2), nor in models with ability peer effects (columns 5-6). Similarly, the coefficients on the region-of-origin variables change only slightly when ability variables are included, with the exception of the coefficient on the fraction of peers from non-German speaking cantons (region 4). Peer ability almost fully explains the positive effects of students from these cantons.

Ability and region-of-origin peer effects display a different relation when explaining the probability of passing the first semester (Table A.6). If both sets of variables are included independently, both sets of variables are significant, at least at the 10%-level (columns 2 and 4). Yet, the effect of peers' region of origin dominates when both sets of variables are included jointly (column 6).

To conclude, Tables A.5 and A.6 highlight two important observations on how ability and region-of-origin peer effects are related. First, ability peer effects operate at least in part independently of region-of-origin peer effects, although region of origin and ability are related. Second, the importance of either ability or region-of-origin peer effects varies across outcome variables.

6 Discussion

This study documents the persistence of initial contacts in a university setting for subsequent outcomes that occur up to three years after the intervention finished. Only a few studies so far investigate whether peer effects in higher education outcomes persist beyond the first year. Sacerdote (2001) finds no persistent effects of roommates' academic ability on students' outcomes after the freshmen year. 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. The present study shows that peer groups based on a one-week intervention can generate peer effects on GPA that persist throughout the first year of college; it is, however, unclear whether the effects persist until graduation. According to the survey results in this study, the formation of friendships and study partnerships are an important channel for academic peer effects, as also suggested by Carrell et al. (2013). Initial friendships, in turn, last beyond the first year.

With respect to the effects of the gender peer group composition, this paper is closest to the study by Oosterbeek and van Ewijk (2014), who investigate the effects of the gender composition of first-year undergraduate working groups. The authors do not find any persistent gender effects. This finding is supported in the present setting. The positive effects of a higher share of females that were established for younger school children (Hoxby, 2000, Lavy and Schlosser, 2011) thus do not seem to translate into such higher education settings.

To the best of my knowledge, this is the first paper to causally investigate the performance effects from peer group composition by region of origin in a higher education setting. It is closest to a number of studies that investigate peer effects from immigrant classmates at the elementary, middle, or high school level. These studies find mixed results that are strongly context dependent. Some studies report that the presence of immigrant children or students who are learners of the native language can hamper the performance of other children in the classroom (Betts, 1998, Gould et al., 2009, Jensen and Rasmussen, 2011, Brunello and Rocco, 2013), while other studies report positive (Hunt, 2012, Bui, 2011), mixed (Friesen and Krauth, 2011), or rather neutral effects (Geay et al., 2013, Ohinata and Van Ours, 2013). In the present study, the peer impact of students from outside Switzerland is overall negative, which is in line with the negative findings at lower levels of schooling as reported by Betts (1998), Gould et al. (2009), Jensen and Rasmussen (2011), Brunello and Rocco (2013); the study thus emphasizes that integrating students from different regional backgrounds creates a challenge for education institutions, even in the absence of language barriers.

A further investigation based on imputed ability suggests that ability peer effects play an important role, even in short-term interventions. In particular, groups with both very high and very low levels of average ability perform substantially worse, compared to groups in the mid-range of average ability. This non-linear pattern is most pronounced for low-ability students. The evidence is in line with findings by Feld and Zölitz (2017), who examine ability peer effects based on randomly assigned undergraduate courses; the authors also find an inverse U-shaped pattern of the effects of peer ability on low-achieving students.

The effect sizes found in this study are relatively small and thus comparable with effect sizes in earlier peer effect studies. For example, for students in the bottom 10% of the peer ability distribution, increasing peer ability by one standard deviation corresponds to an increase in first-year GPA by 0.02 standard deviations. The size of this effect is about 25% of the effect found by Carrell et al. (2009) in their study on cohorts of the US Air Force Academy. In the Air Force Academy, an increase in peers' verbal SAT by one standard deviation on average corresponds to an increase in freshmen GPA by 0.08 standard deviations. The effect sizes found by Carrell et al. (2009) may be larger, compared to the peer effects in the present study, as peer groups at the Air Force Academy stay together for a full year. Moreover, the imputed measure of ability used in the present study may only capture a portion of the variation in peer ability that one observes using

SAT scores.[19]

Given the non-linear nature of ability peer effects, this study suggests that the principal of the university should create groups that are mixed (or "integrated", see the terminology as used by Graham et al., 2010), as opposed to creating groups that are segregated by ability. The policy recommendation from the present study thus deviates from the findings by Booij et al. (2017), who recommend ability tracking. The reason for such a deviation is twofold: First, in contrast to the model by Booij et al. (2017), the model used in the present paper does not include the standard deviation of peer ability. Second, Booij et al. (2017) study a classroom-type setting where peers can directly interfere with each others' learning pace. By contrast, in the present study, students can chose with whom to interact after the first week. Thus, it may be worthwhile to facilitate contacts with diverse peers in informal settings early on, but to allow for tracking later on when assigning students to seminars and workgroups.

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-achievers 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.

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.

The main findings of the paper are twofold: First, the mandatory freshmen week has a long-lasting impact on the formation of friendships and study partnerships. Teammates from the first week are likely to form persistent friendships and are likely to select into the same review sections.

[19]It is important to remember that orientation week peers account only for a small portion of a student's actual friends. Thus, the effect of friends' ability on student outcomes may be much larger than the effect of team members' ability on student outcomes.

Moreover, teammates from the first week are over-represented among friends, even years into their degree.

Second, the composition of the initial peer group has persistent academic impacts. In particular, I find adverse effects of both low- and high-achieving peers on low-achieving students. Overall, an allocation that mixes groups in terms of ability and other student characteristics may benefit average student achievement and minimize adverse effects on low-ability students.

References

BACK, M. D., S. C. SCHMUKLE, AND B. EGLOFF (2008): "Becoming Friends by Chance," *Psychological Science*, 19, 439–440.

BETTS, J. R. (1998): "Educational crowding out: do immigrants affect the educational attainment of American minorities?" in *Help or Hindrance: The Economic Implications of Immigration for African Americans*, ed. by D. Hamermesh and F. Bean, New York: Russell Sage Foundation, 253–81.

BLUNDELL, R. W. AND J. L. POWELL (2004): "Endogeneity in Semiparametric Binary Response Models," *The Review of Economic Studies*, 71, 655–679.

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.

BRUNELLO, G. AND L. ROCCO (2013): "The effect of immigration on the school performance of natives: Cross country evidence using PISA test scores," *Economics of Education Review*, 32, 234–246.

BUI, S. A. (2011): "How do limited English proficient students affect each others' educational outcomes? Evidence from student panel data," Manuscript.

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.

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.

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

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

Friesen, J. and B. Krauth (2011): "Ethnic enclaves in the classroom," *Labour Economics*, 18, 656–663.

Gall, T. L., D. R. Evans, and S. Bellerose (2000): "Transition to first-year university: Patterns of change in adjustment across life domains and time," *Journal of Social and Clinical Psychology*, 19, 544–567.

Geay, C., S. McNally, and S. Telhaj (2013): "Non-native Speakers of English in the Classroom: What Are the Effects on Pupil Performance?" *The Economic Journal*, 123, 281–307.

Gould, E. D., V. Lavy, and M. Daniele Paserman (2009): "Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi-Experimental Evidence," *The Economic Journal*, 119, 1243–1269.

Graham, B. S. (2011): "Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers," in *Handbook of Social Economics*, ed. by J. Benhabib, A. Bisin, and M. O. Jackson, North-Holland, vol. 1, chap. 19, 965–1052.

Graham, B. S., G. W. Imbens, and G. Ridder (2010): "Measuring the Effects of Segregation in the Presence of Social Spillovers: A Nonparametric Approach," Working Paper 16499, National Bureau of Economic Research.

HOXBY, C. (2000): "Peer Effects in the Classroom: Learning from Gender and Race Variation," Working Paper 7867, National Bureau of Economic Research.

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

HUNT, J. (2012): "The Impact of Immigration on the Educational Attainment of Natives," Working Paper 18047, National Bureau of Economic Research.

JENSEN, P. AND A. W. RASMUSSEN (2011): "The effect of immigrant concentration in schools on native and immigrant children's reading and math skills," *Economics of Education Review*, 30, 1503–1515, Special Issue: Economic Returns to Education.

LAVY, V. AND A. SCHLOSSER (2011): "Mechanisms and Impacts of Gender Peer Effects at School," *American Economic Journal: Applied Economics*, 3, 1–33.

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.

OECD (2009): *Education at a Glance 2009: OECD Indicators*, OECD Publishing, Paris.

OHINATA, A. AND J. C. VAN OURS (2013): "How immigrant children affect the academic achievement of native Dutch children," *The Economic Journal*, 123, 308–331.

OOSTERBEEK, H. AND R. VAN EWIJK (2014): "Gender peer effects in university: Evidence from a randomized experiment," *Economics of Education Review*, 38, 51–63.

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 (2016): "Nudging Generosity: Choice Architecture and Cognitive Factors in Charitable Giving," USC-INET Research Paper No. 16-26.

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. (2014): "On the Nature of Social Interactions in Education: An Explanation for Recent Puzzling Evidence," Manuscript.

TINTO, V. (1975): "Dropout from Higher Education: A Theoretical Synthesis of Recent Research," *Review of Educational Research*, 45, pp. 89–125.

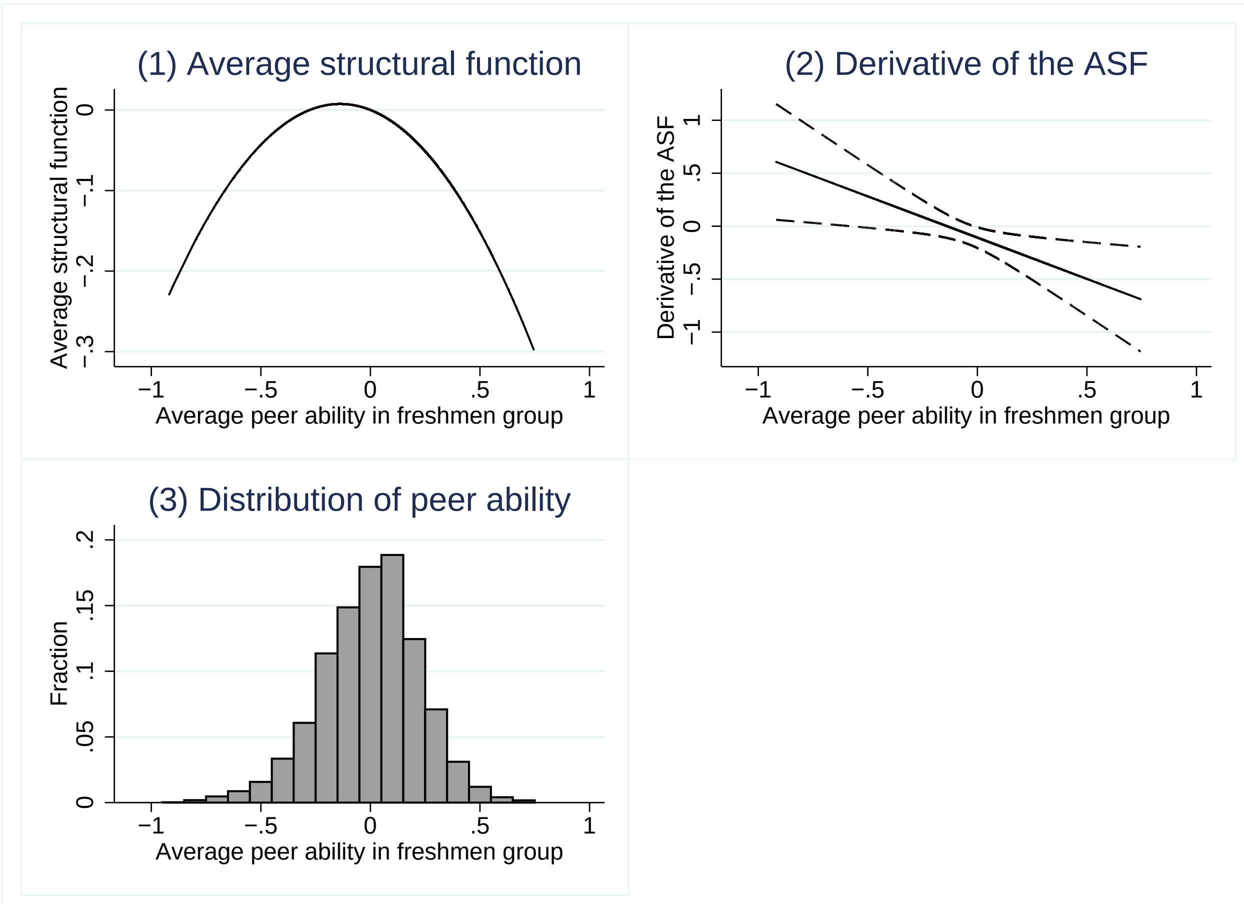
Appendix

Figure A.1: Schedule of a typical freshmen group

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

Note: Freshmen week schedule of a typical freshmen 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. Source: Own illustration based on a schedule of a typical freshmen group by the St. Gallen administration.

Figure A.2: Effect of peer ability (imputed) on first-year GPA



Note: The figure presents the Average Structural Function (ASF) of the relation between imputed peer ability and first-year GPA (Panel 1), and its derivative (Panel 2) across the support of imputed ability. For details on the imputation, see Section 5.4. The functions are based on a quadratic model (see Table 4). The dashed lines in Panel 2 represent 95% confidence intervals, based on analytic standard errors. Panel 3 shows the distribution of peer ability in the sample (7,953 observations). Source: Own calculations based on administrative data from the University of St. Gallen.

Table A.1: Test for random assignment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent variables: Group composition						
	Peers' gender	Peers' region of origin					
	Fraction:	Fraction:					
	Female	Region 1	Region 2	Region 3	Region 4	Region 5	Foreign
Individual characteristics							
Age (years)	0.001	-0.000	-0.000	0.000	-0.000	0.000	-0.000
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Region 1	-0.008*	0.001	0.003	0.001	0.001	0.002	-0.001
	(0.004)	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Region 3	0.002	-0.002	-0.002	0.002	-0.003	0.003	0.001
	(0.004)	(0.006)	(0.004)	(0.002)	(0.002)	(0.003)	(0.003)
Region 4	-0.004	-0.007	0.002	-0.001	0.005	0.005	-0.000
	(0.007)	(0.009)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)
Region 5	-0.001	-0.002	-0.007	0.010***	-0.004	0.004	0.005
	(0.006)	(0.007)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)
Non-Swiss nationaliy	-0.008	0.010*	0.000	0.001	-0.002	-0.002	0.002
	(0.005)	(0.006)	(0.004)	(0.003)	(0.002)	(0.003)	(0.003)
Non-German mother	0.003	-0.004	-0.000	-0.002	0.001	0.001	0.000
tongue	(0.006)	(0.007)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)
Stratifying variables							
Female	-0.003	-0.002	0.001	0.000	0.001	0.003***	-0.050***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Foreign HS degree	0.018***	0.003	0.006	0.004	0.005	-0.036***	0.000
	(0.006)	(0.007)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.050	0.028	0.044	0.041	0.014	0.121	0.148
Number of students	8,073	8,073	8,073	8,073	8,073	8,073	8,073
Number of groups	526	526	526	526	526	526	526

Note: The table presents coefficients from OLS regressions of treatment variables (group composition variables) on student characteristics. Controls include cohort dummies and the size of the freshmen week team. Analytic standard errors, clustered at the group level, are in parentheses. Source: Own calculations based on administrative data from the University of St. Gallen.
*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	
Region 1 (close to the university)	-0.050*
	(0.028)
Region 3 (mixed-language canton)	-0.047
	(0.038)
Region 4 (non-German speaking canton)	-0.275***
	(0.066)
Region 5 (Swiss, HS degree abroad)	-0.373***
	(0.054)
Foreign (foreign, HS degree foreign)	0.755***
	(0.052)
Further characteristics	
Non-German mother tongue	-0.277***
	(0.049)
Foreign 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. Source: Own calculations based on administrative data from the University of St. Gallen.
*p < 0.10, **p < 0.05, ***p < 0.01.

Table A.3: Effect heterogeneity by gender

	(1)	(2)
	Dependent variable: GPA	
Sample	Female	Male
Peers' gender		
Fraction female	0.189	-0.058
	(0.258)	(0.427)
Peers' region of origin		
Fraction region 1 (close to St. Gallen)	0.173	-0.243
	(0.199)	(0.039)
Fraction region 3 (mixed-language canton)	-0.121	0.079
	(0.315)	(0.294)
Fraction region 4 (non-German speaking canton)	0.734**	-0.118
	(0.012)	(0.314)
Fraction region 5 (Swiss, HS degree abroad)	-0.250	-0.396*
	(0.223)	(0.051)
Fraction foreign (foreign, HS degree foreign)	-0.240	-0.279*
	(0.204)	(0.085)
Obs.	2,525	5,428

Note: The table presents average marginal effects (AMEs) from regressions of student first-year GPA on group composition variables. The p-value is based on randomization inference (1,000 random draws from the randomization distribution). The regressions are based on linear models and estimated using OLS. All regressions include cohort dummies and controls (gender, region of origin, admission test indicator, age, group size). Source: Own calculations based on administrative data from the University of St. Gallen.
*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A.4: Effect heterogeneity by region of origin

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: GPA					
Sample	Region 1	Region 2	Region 3	Region 4	Region 5	Foreign
Peers' gender						
Fraction female	-0.236	-0.096	0.381	0.260	-0.011	0.445
	(0.269)	(0.303)	(0.223)	(0.427)	(0.472)	(0.111)
Peers' region of origin						
Fraction region 1 (close to St. Gallen)	0.067	-0.087	0.383	-0.841*	-0.294	-0.347*
	(0.372)	(0.295)	(0.129)	(0.069)	(0.312)	(0.071)
Fraction region 3 (mixed-language canton)	-0.054	0.244	-0.173	-0.018	-0.904	-0.061
	(0.433)	(0.149)	(0.390)	(0.490)	(0.141)	(0.431)
Fraction region 4 (non-German speaking canton)	0.506	-0.104	0.988*	0.312	-1.060	0.051
	(0.087)	(0.368)	(0.058)	(0.355)	(0.158)	(0.427)
Fraction region 5 (Swiss, HS degree abroad)	-0.421	-0.410	0.212	-0.758	-1.868*	-0.148
	(0.140)	(0.111)	(0.379)	(0.224)	(0.060)	(0.377)
Fraction foreign (foreign, HS degree foreign)	-0.502	-0.072	0.823*	-1.158	-1.007	-0.508
	(0.061)	(0.384)	(0.059)	(0.102)	(0.109)	(0.110)
Obs.	1,979	2,861	853	463	363	1,434

Note: The table presents average marginal effects (AMEs) from regressions of student first-year GPA on group composition variables. The p-value is based on randomization inference (1,000 random draws from the randomization distribution). The coefficients are estimated using OLS. All regressions include cohort dummies and controls (gender, region of origin, admission test indicator, age, group size). Source: Own calculations based on administrative data from the University of St. Gallen.
*p < 0.10, **p < 0.05, ***p < 0.01.

Table A.5: Robustness and channels (GPA)

	(1)	(2)	(3)	(4)	(5)
	Dependent variable: GPA				
Peers' gender					
Fraction female	0.034	-0.394			0.006
	(0.143)	(0.709)			(0.144)
Fraction female (squared)		0.688			
		(1.109)			
Peers' region of origin					
Fraction region 1	-0.115	-0.068			-0.134
	(0.101)	(0.316)			(0.102)
Fraction region 1 (squared)		-0.074			
		(0.557)			
Fraction region 3	0.008	0.274			-0.010
	(0.135)	(0.332)			(0.134)
Fraction region 3 (squared)		-0.972			
		(1.143)			
Fraction region 4	0.153	0.527			0.000
	(0.166)	(0.395)			(0.197)
Fraction region 4 (squared)		-1.994			
		(1.773)			
Fraction region 5	-0.367*	0.277			-0.477**
	(0.201)	(0.534)			(0.208)
Fraction region 5 (squared)		-4.269			
		(3.475)			
Fraction foreign	-0.244	-0.112			-0.114
	(0.157)	(0.499)			(0.181)
Fraction foreign (squared)		-0.383			
		(1.247)			
Peer ability					
Avg. peer ability			-0.076	-0.101**	-0.096
			(0.049)	(0.050)	(0.062)
Avg. peer ability (squared)				-0.382**	
				(0.155)	
Controls	Yes	Yes	Yes	Yes	Yes
Obs.	7,953	7,953	7,953	7,953	7,953
Adj. R-squared	0.066	0.065	0.066	0.066	0.066
p-value of F-test					
F-test: Peers' gender	0.811	0.798	-	-	0.966
F-test: Peers' region of origin	0.137	0.312	-	-	0.186
F-test: Peers' ability	-	-	0.125	0.020	0.124
Joint F-test of all peer variables	0.211	0.461	0.125	0.020	0.118

Note: The table presents coefficients from regressions of student first-year GPA on group composition variables. The models are estimated using OLS. The ability measure is imputed based on leave-own-year-out regressions of first-year GPA on student characteristics (gender, region of origin, admission test indicator, mother tongue, nationality, age, age squared, cohort dummies). See Section 5.4 for details on the computation of the ability measure. Analytic standard errors, clustered at the group level, are in parentheses. All regressions control for cohort dummies and group size. Source: Own calculations based on administrative data from the University of St. Gallen. *p < 0.10, **p < 0.05, ***p < 0.01.

Table A.6: Robustness and channels (first semester passed)

	(1)	(2)	(3)	(4)	(5)
	Dependent variable: First semester passed				
Peers' gender					
Fraction female	-0.061	-1.024			-0.077
	(0.234)	(1.232)			(0.237)
Fraction female (squared)		1.555			
		(1.910)			
Peers' region of origin					
Fraction region 1	-0.095	0.157			-0.105
	(0.163)	(0.555)			(0.164)
Fraction region 1 (squared)		-0.456			
		(1.020)			
Fraction region 3	-0.155	0.465			-0.164
	(0.220)	(0.541)			(0.218)
Fraction region 3 (squared)		-2.283			
		(1.847)			
Fraction region 4	0.663**	1.196*			0.582*
	(0.281)	(0.633)			(0.327)
Fraction region 4 (squared)		-2.866			
		(3.035)			
Fraction region 5	-0.614*	0.641			-0.672**
	(0.315)	(0.742)			(0.328)
Fraction region 5 (squared)		-8.093*			
		(4.278)			
Fraction foreign	-0.342	-0.723			-0.275
	(0.256)	(0.891)			(0.300)
Fraction foreign (squared)		0.931			
		(2.177)			
Peer ability					
Avg. peer ability			-0.102	-0.121*	-0.050
			(0.077)	(0.073)	(0.101)
Avg. peer ability (squared)				-0.340	
				(0.225)	
Controls	Yes	Yes	Yes	Yes	Yes
Obs.	8,073	8,073	8,073	8,073	8,073
Pseudo R-squared	0.054	0.055	0.052	0.053	0.054
p-value of F-test					
F-test: Peers' gender	0.793	0.708	-	-	0.746
F-test: Peers' region of origin	0.033	0.019	-	-	0.075
F-test: Peers' ability	-	-	0.184	0.083	0.622
Joint F-test of all peer variables	0.056	0.037	0.184	0.083	0.072

Note: The table presents coefficients from regressions of an indicator variable (whether a student passed the first semester) on group composition variables. The regressions are based on probit models and estimated using MLE. The ability measure is imputed based on leave-own-year-out regressions of first-year GPA on student characteristics (gender, region of origin, admission test indicator, mother tongue, nationality, age, age squared, cohort dummies). See Section 5.4 for details on the computation of the ability measure. Analytic standard errors, clustered at the group level, are in parentheses. All regressions control for cohort dummies and group size. Source: Own calculations based on administrative data from the University of St. Gallen.
*p < 0.10, **p < 0.05, ***p < 0.01.