Classroom or Pub – Where are Persistent Peer Relationships between University Students Formed?*

Thomas Fischer and Johannes Rode†

October 15, 2019

This paper discusses the formation of peers in higher education using a unique data set of industrial engineering students. For identification, we exploit the random assignment of students into groups and student performance before students met. We compare two different settings for potential peer formation: a voluntary freshman orientation week organized by the students' union and a mandatory group work course. It is only in the case of the group work course that we report persistent impacts on subsequent academic achievement. In line with our theoretical reasoning, peer effects exist between groups of two students who were already similar before.

JEL classification: I21 – I23 – D85

Keywords: peer effects – higher education – social network formation – homophily

1. Introduction

First-year bachelor students jump in at the deep end. For most of them it is the first time to live without their parents and far away from their high-school friends. On the one hand, this offers the opportunity to develop as an individual without baggage from

*We would like to thank Petra Thiemann for providing us with detailed comments.

†Fischer: Lund University, Sweden, thomas.fischer@nek.lu.se; Rode: Technische Universität Darmstadt, Germany, rode@vwl.tu-darmstadt.de

the past. On the other hand, students are thrown into a competitive and academically challenging environment without the support of their existing peers.

In this paper, we empirically investigate how students form peer groups that affect their subsequent academic success. It is well known that individuals are influenced by their peers – in particular in education settings.[1] This holds true for (high) school education (e.g. Kang (2007), Burke & Sass (2013)), but also for higher education: Studies have exploited the co-habiting structure of dorms to document peer effects (cf. e.g. Zimmerman (2003), Sacerdote (2001)). Several papers (e.g. Carrell et al. (2009), Carrell et al. (2013), or Lyle (2009)) consider the specific case of military cadets who have a very close interaction with their fellow students.

These findings might be hard to generalize to more traditional university settings. Moreover, little is known about the formation of peer groups, making it hard to induce peer effects in a form of *social engineering*. For example, researchers have tried to impose positive peer effects (especially for otherwise low-performing students) with a mentor system (Carrell et al., 2013). Interestingly, this experiment failed as individuals repelled their assigned partners and matched with more homophilic peers. This result is in line with the successful intervention presented in Oosterbeek et al. (2016), for which students were pre-sorted according to prior ability.

In this paper, we shed light on the formation of peer groups that might hinder active interventions. We consider detailed information in the competitive environment of industrial engineering students at *Technische Universität Darmstadt*, one of the leading universities specialized in engineering subjects in Germany. Every year many students are admitted resulting in large introductory classes with several hundred students with little interaction between students. Furthermore, class attendance is not mandatory, the curriculum does not feature group work, and students usually do not live on campus. All of this contributes to a very anonymous structure differing considerably from the settings usually considered in the literature in which the exposure to other students comes by cohabiting or due to small and interactive class rooms. In our setting, the university students are, however, more narrowly exposed to smaller groups of other individuals in two university-related activities providing them with new connections to form (potential) peers. These exposures are not directly designed for this research, but represent a quasi natural experiment lasting over several years and cohorts of students. More

[1] Epple & Romano (2011) and Sacerdote (2011) summarize the literature.

precisely and as illustrated in Figure 1, we exploit the random assignment of students into groups and information on student performance before students have met.

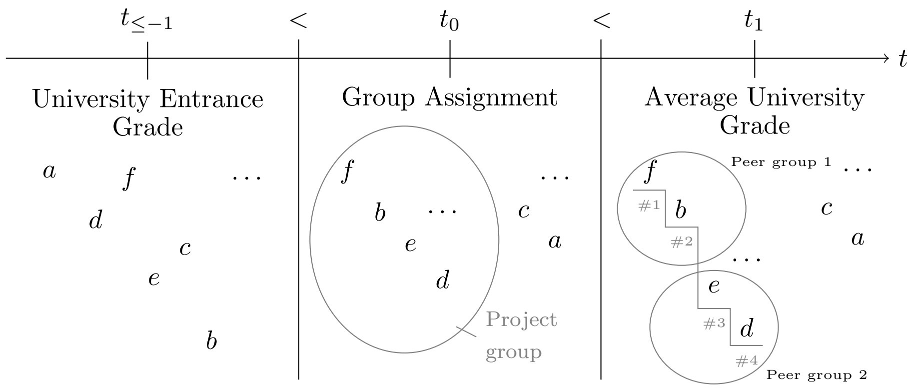

Figure 1: Timeline of events and the next-best peer effect.

Notes: $a, b, c, d, e, f, \cdots$ are students. At $t_{\leq -1}$, students receive the University entrance grade from high school. At this point in time, students cannot influence each other since they have not met their group assignment peers. At t_0, students are randomly allocated to project groups. Students f, b, e, d may, for example, be randomly allocated to the same group. At t_1, the next-best peer effect appears: students who are similar in terms of pre-university ability form, e.g., study groups and influence each other (in our case student f may influence b and student d may influence e).

In the middle of their bachelor studies, students are required to take part in an one-week intensive group work course, in which they are randomly assigned into a group, allowing them to form new connections. Given the academic nature of the event, it is also possible to observe the academic abilities and work ethics of their fellow students. We document that the group as a whole (as measured by the group average minus the respective individual, Manski (1993)) does not affect the subsequent academic outcome of the individual.

The peer exposure is only prevalent for a short period of time. After one intensive week of interaction, students can part ways. Despite this fact, we document a more narrow effect between a pair of students which only sets in after establishing connections in the group course. In our setup, we document a *next-best* peer effect, i.e. peer effects exist between individuals who are already similar in terms of academic strength ex-ante. In this setting, peers are not superimposed by some exogenous rule, but are self-selected from the exogenously given (random) set of connections. This is in line with Carrell et al. (2013). In their work, they try to exploit the documented peer effect in order to design optimal groups. Carrell et al. (2013) show that this experiment failed and individuals avoided the assigned peers and matched up with individuals already similar to themselves. This result does not emerge when conducting a placebo test in which we assign students of similar ability to each other that, however, have no existing connection.

Despite the group work only lasting for one week, the next-best peer effect is still measurable even several years after the group work took place. We provide a battery of robustness checks to address concerns with respect to randomness of the assignment and the measurement of student achievement. Moreover, we document that male students are more prone to be influenced by peers and that the identified effect persists over time.

We contrast our analysis with another random group assignment. At the very beginning of bachelor studies, students are randomly matched in small groups in a so-called *orientation week*. This informal social event organized by the students' union aims at establishing early contacts between students. In this setting, individuals might become *friends by chance* (Back et al., 2008). As suggested by our empirical analysis, there is, however, no significant peer influence on subsequent academic outcome.

Our results are in line with a simple theoretical reasoning, which we develop. We formally show that in order to maximize the individual academic outcome, it is helpful to match with individuals of the same type. While it might be attractive for those with low ability to spend a lot of time in groups of high ability students, the latter have no incentive to participate in this match. Thus, in a bargaining equilibrium, individuals of homogenous type team up. It is also important to point out that this is not due to some exogenous preference for homophily as assumed in other papers (e.g. Brady et al. (2016)). The model furthermore shows that – even without a socializing preference – individuals are always better off in terms of academic achievement when working in groups than in autarky.

The failure of the social setting in creating peer effects is likely to be found in the informal nature of this event. In line with the theoretical reasoning, there might be a mismatching in study groups due to the noisy signal perceived about academic strength in the social setting, leading them to discontinue, and thus not create a peer effect. In contrast, but in line with that reasoning, Thiemann (2018) also discusses an orientation week that involves students working on a formal case study and thereby creates a long-lasting peer impact.

Our findings are relevant for the design of degree programs. Group work can indeed have a positive impact on student achievement. According to our results, directors of degree programs should make sure that group formation occurs in an academic context.

The remainder of this paper is structured as follows. Section 2 proposes our peer measure prevailing between individuals of already similar type. We motivate it with a simple theoretical model (cf. Section 2.1) and also discuss how to address empirical identification challenges in this setting (cf. Section 2.2). We introduce our specific institutional setting in Section 3 and discuss our findings in Section 4. The final Section concludes.

2. Theoretical background

We explore two group events to identify individuals who know each other. This set, however, only forms connections. In contrast to (high) school students that are subject to enforced interaction in the classroom setting, the university students have the option to pick their own peers. In the following Section, we develop a simple theoretical model in which the formation of peers is an endogenous decision made in order to maximize study outcome. In Section 2.2, we show the implications of this setup for the empirical estimation.

2.1. A simple model

We consider a scenario in which individuals are free to decide who their peers are and how much exposure time τ they spend together. The simple model not only shows that individuals connect to other individuals comparable to themselves, but also argues that all individuals gain from group work. Note that this does not result from an exogenous assumed homophilic preference as in Brady et al. (2016), but originates endogenously in heterogeneous skills.

In essence, we try to explain the academic outcome of exams which take the form of an average university grade G_i for individual i. We assume that the latter is determined by two key factors: (i) exogenous prior knowledge I_i and (ii) time-consuming repetition. While individuals repeat on their own, new knowledge can only be transferred by connecting to other individuals by forming study groups. Essentially, the study group contains a positive externality.

Assume that academic outcomes test a certain information set I, which we normalize to $I \equiv 1$. Due to exogenous heterogeneous skills or to heterogeneous participation rates in the lectures each individual i only has access to a subset of the total necessary knowledge $0 < I_i < 1$.

Secondly, the skills have to be repeated to be retrievable for examinations. In total, all individuals have a finite and identical amount of time T which they can employ for practicing or to attend group work.

In autarky (label A), all time T – which we normalize to $T \equiv 1$ – is employed to practice the given knowledge I_i. The education outcome is given by:

$$G_i(A) = \alpha I_i. \tag{1}$$

The measure $0 < \alpha < 1$ captures the strength of translation between the information measurement I – as a measure of individual ability – into an education outcome in a simple linear model.

Now, we want to contrast this result with a world in which group work is possible. Rather than living in autarky, individuals now have access to external connections with whom they can form learning groups. Due to the individual finite knowledge $0 < I_i < 1$ it is useful to connect to other individuals.

We assume that learning groups only constitute of two individuals. This can be rationalized by the assumption that the group outcome is determined by the *weakest link*. More formally and following the general model structure of Bénabou (1996), we assume that the group outcome is aggregated by a Constant Elasticity of Substitution (CES) measure with an elasticity of substitution ϵ. For the special case with $\epsilon \to \infty$ (Leontief production function) the overall output is determined by the weakest link. An individual seeking to maximize the group outcome would only admit the best individuals and would keep the group size as small as possible, finally leading to pairs.[2] In the empirical framework this will also help to more clearly demonstrate the peer effect. With a larger group the pattern of influences would be less clear-cut.

We consider a simple specification with individual studying subject to decreasing returns to scale. By repeating their given skills (e.g. solving specific math problems or repeating the context of text books), individuals become better at it, yet at a decreasing rate. Moreover and even more importantly, they will not gain new knowledge. The transfer of knowledge itself depends on the exposure from the other individual (indexed with

[2] Formally, the individual optimizes $\max \min_{i=1}^{n}\{I_i\}$, implying $n = 1$ and $\max_i\{I_i\}$.

$-i$) and on the exposure time τ. The linear specification is in line with the estimation model. As such, we specify the overall education grade production function:

$$0 < G_i = (\alpha I_i + \beta I_{-i}\tau_i)(1-\tau_i)^\gamma < 1, \tag{2}$$

with $0 < \gamma < 1$ capturing negative scale effects for individual studying, also guaranteeing that the goal function G_i is concave in the control τ_i, which goes along with a global maximum. Jointly with the scale-free property of the group activity, the overall education production function exhibits positive scale effects ($\gamma + 1 > 1$). The parameter $0 < \beta < 1$ captures the strength of a linear peer effect given the exposure time τ_i. It is also easy to see that the autarky case (eq. 1) is nested for $\tau_i = 0$. Individuals choose their participation time in the study group τ_i in order to maximize their grade outcome G_i. The first-order condition implies:

$$\tau_i^* = \frac{1}{1+\gamma} - \frac{\alpha\gamma I_i}{I_{-i}(1+\gamma)\beta}. \tag{3}$$

Not surprisingly, the total group participation depends positively on the peers information set I_{-i} and negatively on the own information endowment I_i. Individuals with high knowledge I are attractive partners, who are able to transfer a lot of knowledge. On the other hand, they are less inclined to spend a lot of time in group work. Consider another individual (index $-i$) with an optimal group work supply:

$$\tau_{-i}^* = \frac{1}{1+\gamma} - \frac{\alpha\gamma I_{-i}}{I_i(1+\gamma)\beta}. \tag{4}$$

As the study group only exists if both individuals are present, the matched pair spends the same amount of time in group work. Formally, the matching condition is $\tau^* = \tau_i^* = \tau_{-i}^*$. From this condition it follows immediately that matching only happens if $I_i = I_{-i}$. As such, individuals with information sets of identical magnitude do match. While low ability students would like to spend more time in the study group, their higher ability counterpart would not be willing to supply this high amount of time. For an efficient matching to happen, both should be of identical magnitude. As mentioned before, this result does not follow from some endogenously assumed preference for homophily as in Brady et al. (2016), but from an equilibrium bargaining idea.

Given this symmetry in the matching, the optimal group participation (eq. 4) boils down to:

$$0 < \tau^* = \frac{1}{1+\gamma} - \frac{\alpha}{\beta}\frac{\gamma}{1+\gamma} < 1, \tag{5}$$

and is identical for all individuals regardless of their specific level of I_i. The factor $\frac{\alpha}{\beta}$ captures the relative strength of the peer effect. A lower ratio, indicating a stronger peer effect, contributes to a higher group work participation τ^*. If the effects of individual learning ability α and peer effects β are of identical magnitude ($\alpha = \beta$) optimal group participation yields $\tau = \frac{1-\gamma}{1+\gamma} < 1$ only depending negatively on the scaling γ of the individual studying effort. Inserting this result in the overall production function, we find an education output under the presence of peers (label P):

$$G_i(P) = I_i \left(\frac{\alpha+\beta}{1+\gamma}\right)\left(\frac{(\alpha+\beta)\gamma}{(1+\gamma)\beta}\right)^{\gamma}. \tag{6}$$

The multiplier of the peer effect is given by:

$$\frac{G_i(P)}{G_i(A)} = \left(\frac{\alpha+\beta}{\alpha(1+\gamma)}\right)\left(\frac{\gamma(\beta+\alpha)}{(1+\gamma)\beta}\right)^{\gamma} \tag{7}$$

For the case of $\alpha = \beta$, one can easily show that this multiplier is larger than one for all $\gamma < 1$. I.e., the multiplier is larger than one as long as negative scale effects from the non-group activity of training exist.

The key equation (eq. 4) is derived under the assumption that individuals have both perfect knowledge about their own ability and the ability of their counterparts. At least, the latter is a questionable assumption. In the core empirical investigation of this paper, we investigate two settings in which individuals meet: a classroom setting and a social gathering. While in the former setting individuals clearly showcase their ability in and effort for university activities, the signal sent in the latter is substantially more noisy. As a result, there is a higher likelihood of mismatching. These mismatches are revealed under once the individuals form their learning groups potentially leading to the groups being discontinued. In line with that idea, we find no significant peer effect in the social setting in our empirical investigation.

2.2. Estimation model

The empirical literature usually measures the peer effect for an individual by the mean overall outcome of all individuals (apart from the very individual to be explained). The identification challenges which emerge in this setup are considered, e.g. in Manski (1993) or in more general in Blume et al. (2011). While this setup is well suited to consider primary education taking place in a classroom setting, it does not fit our framework, in which individuals can avoid others. In fact, the argument in Manski (1993) relies on a *law-of-large-numbers* argument. Our theory suggests the diametrical extreme with only two individuals exerting influence on each other. Epple & Romano (2011, Section 3.2) consider a similar setup. Assume we want to explain the educational outcome G of an individual i by means of her skills S_i and factors depending on her peer (indicated by $-i$) implying the following regression equation:

$$G_i = \phi_0 + \phi_1 S_i + \phi_2 G_{-i} + \phi_3 S_{-i} + \epsilon_i, \tag{8}$$

respectively for individual $-i$:

$$G_{-i} = \phi_0 + \phi_1 S_{-i} + \phi_2 G_i + \phi_3 S_i + \epsilon_{-i}, \tag{9}$$

with $\epsilon \sim N(0, \sigma)$. Of course, we cannot observe the level of skills S directly. Under the label S, however, we capture a complete vector of variables that determine the level of skills. It is important to emphasize that the factors captured in S are observed before the peer relationship was formed. This in particular holds true for factors which do not vary in time (e.g. foreign origin). Combining both equations to cancel out factors emerging jointly in time (i.e. the respective education outcomes G) we get:[3]

$$G_i = a + bS_i + cS_{-i} + \eta_i, \tag{10}$$

[3]We also have $\eta_i = \frac{\epsilon_i + \epsilon_{-i}\phi_2}{1-\phi_2^2}$ which under $Corr(\epsilon_i, \epsilon_{-i}) = 0$ implies $\eta \sim N\left(0, \frac{\sqrt{1+\phi_2^2}}{1-\phi_2^2}\sigma\right)$.

with:

$$
\begin{aligned}
a &= \frac{\phi_0}{1-\phi_2}, \\
b &= \frac{\phi_1 + \phi_2\phi_3}{1-\phi_2^2}, \\
c &= \frac{\phi_1\phi_2 + \phi_3}{1-\phi_2^2},
\end{aligned}
\tag{11}
$$

for which b captures the effect of own ability (in a peer setting) and c measures the peer effect. In fact, the latter captures both the endogenous (captured in ϕ_2) and the exogenous peer effect (as considered by ϕ_3). The endogenous peer effect ϕ_2 is exerted by actions taken simultaneously, whereas the exogenous peer effect follows from the pure presence of a peer.

Epple & Romano (2011) suggest that there might be some unobserved correlation $\phi_4 = corr(\epsilon_i, S_{-i})$. The presence of $\phi_4 \neq 0$ implies that individuals self-select a partner based on measures that are not observed by the researcher. As such, measuring $c \neq 0$ does not necessarily imply the presence of a peer effect. Finally, the measure a captures a context effect, i.e. a factor that all individuals were superimposed jointly (e.g., the fact that an exam was particularly hard at a certain point in time).

For the special case of $S_i = S_{-i}$ – as suggested in the theoretical model – we would have $\bar{b} = \bar{c} = b + c = \frac{(1+\phi_2)(\phi_1+\phi_3)}{1-\phi_2^2} = \frac{\phi_1+\phi_3}{1-\phi_2}$. For $0 < \phi_2 < 1$ this captures the reflection effect. In autarky $\phi_2 = \phi_3 = 0$ the individual ability S_i would transfer in a linear manner into a grade G_i by means of a coefficient ϕ_1. In the presence of peers, the multiplier $\frac{1}{1-\phi_2}$ reveals the peer effect.[4] The same logic holds true for the exogenous peer effect ϕ_3.

Of course, the assumption of $S_i = S_{-i}$ will not hold exactly in the empirical evaluation. The theory suggests that individuals match with a similar peer. Given the finite amount of connections in the data, we match them with a *next-best* peer – i.e. an individual who exhibits the lowest absolute deviation from themselves in terms of prior ability S. With the usual law-of-large-numbers argument this imprecise matching will still cancel out in the aggregate and thus only appear in the overall error term ϵ_i.

[4]Formally, this is easy to see by the standard geometric series logic $\sum_{i=0}^{\infty} \phi_2^i = \frac{1}{1-\phi_2}$.

In autarky, i.e. in the absence of both peer effects ($\phi_2 = \phi_3 = 0$), we would have $b = \phi_1$ and $c = 0$ measuring an overall zero peer effect. The expected gain from peer exposure is thus:[5]

$$E[G(P) - G(A)] = \frac{\phi_0 \phi_2}{1 - \phi_2} + \frac{\phi_1 \phi_2 + \phi_3}{1 - \phi_2} S_i > 0. \tag{12}$$

We compare this regression model with a scenario based on a mean peer effect, which is usually employed in the literature. In this case, an individual i is part of a peer group k for which the peer measure is given by $\bar{X}_{i,k} = \sum_{l \neq i} X_{l,k}$.[6] A general regression is specified as follows:

$$G_{i,k} = \tilde{\phi}_0 + \tilde{\phi}_1 S_i + \tilde{\phi}_2 \bar{G}_{i,k} + \tilde{\phi}_3 \bar{S}_{i,k} + \tilde{\epsilon}_i. \tag{13}$$

Taking the definition of the peer measure, we find the following overall reduced specification:

$$G_{i,k} = \tilde{a} + \tilde{b} S_i + \tilde{c} \bar{S}_{i,k} + \tilde{\epsilon}_i, \tag{14}$$

with:

$$\begin{aligned} \tilde{a} &= \frac{\tilde{\phi}_0}{1 - \tilde{\phi}_2}, \\ \tilde{b} &= \tilde{\phi}_1, \\ \tilde{c} &= \frac{\tilde{\phi}_1 \tilde{\phi}_2 + \tilde{\phi}_3}{1 - \tilde{\phi}_2}. \end{aligned} \tag{15}$$

As before, the presence of an endogenous peer effect $\phi_2 \neq 0$ complicates the identification. While the contextual peer effect $\tilde{a}$ is the same as in the two-person setting a, the individual ability effect remains unaffected by the reflection problem $\tilde{b} = \tilde{\phi}_1$ due to the assumed law-of-large-numbers property. Finally, while the peer effect $\tilde{c}$ as compared to c is of the same structure in the nominator, it differs in the denominator and is expected to be lower for the mean-peer effect ($\tilde{c} < c$).[7] In fact, the social multiplier[8] is lower than in the two-person case $\tilde{\phi}_2/(1 - \tilde{\phi}_2) > 1/(1 - \phi_2)$ as here the initial ability is first

[5] The sign of the effect is under the reasonable assumptions $\phi_i > 0 \, \forall \, i$ and $\phi_2 < 1$.

[6] In order to differentiate from the two person case, which we discussed so far, we now employ the tilde operator.

[7] Formally, this requires the established assumptions $\phi_2 = \tilde{\phi}_2 < 1$.

[8] It follows from the geometric series starting with $i = 1$: $\sum_{i=1}^{\infty} \tilde{\phi}_2^i = \frac{\tilde{\phi}_2}{1 - \tilde{\phi}_2}$

captured in the aggregate mean and then perpetually self-reflected. In contrast, in the two-person setting with perfect matching there is direct self-reflection.

In this case, the expected gain from peer exposure amounts to:

$$E[G(P) - G(A)] = \frac{\tilde{\phi}_0 \tilde{\phi}_2}{1 - \tilde{\phi}_2} + \frac{\tilde{\phi}_1 \tilde{\phi}_2 + \tilde{\phi}_3}{1 - \tilde{\phi}_2} \bar{S}_{i,k} = \frac{\tilde{\phi}_0 \tilde{\phi}_2}{1 - \tilde{\phi}_2} + \tilde{c} \bar{S}_{i,k} > 0. \tag{16}$$

If the estimated coefficients are the same $\left(\tilde{\phi}_i = \phi_i \, \forall \, i\right)$, this would imply a gain from peer exposure comparable to the one put forward in equation 12. The subtle – but important – difference in the first setting is that the peer multiplier $\tilde{c}$ applies to the individual level whereas in the second it applies to the average skill level $\bar{S}$. As such, above average individuals would prefer a one-by-one matching, whereas those below the mean in terms of given skills S profit more from a general group exposure.

3. Institutional background and data

We employ exclusive data from students at the department of *law and economics* at Technische Universität Darmstadt, Germany. Technische Universität Darmstadt is a member of TU9, the alliance of nine important technical universities in Germany and is specialized in engineering programs. We capture students enrolled in the field of *Wirtschaftsingeniuerwesen* or *Wirtschaftsinformatik* (engl. industrial engineering or business information systems). This is the largest sub-field within the university and the university has been ranked among the top 3 universities in this field in Germany in recent years (WirtschaftsWoche, 2014, 2016). The field of studies is split into two parts. One half consists of courses in law, economics, and business administration, while the second half are engineering classes. For the latter, students can choose (ranked by popularity in our sample) mechanical engineering, electrical engineering, civil engineering, and computer science.[9] Table 1 shows the shares for the students' technical major. Given the excellent reputation of the university, the field of study is considered to be rather demanding as manifested in high failure rates in exams and high drop-out rates in general.

Regardless of the chosen major, students study all courses in law, economics, and business together. It is only for the engineering parts – considered to be more difficult, as documented by higher failure rates – that the majors are split. In our sample, approx.

[9] Due to a change in curriculum, we have only a few students with computer science as a major.

Table 1: Shares by (engineering) major and gender.

Engineering Major	Male	Female	Total
Electrical Engineering (in %)	12.7	11.7	12.6
Civil Engineering (in %)	8.7	30.6	11.9
Computer Science (in %)	2.5	1.8	2.4
Mechanical Engineering (in %)	76.1	55.9	73.1
Total (in %)	100.0	100.0	100.0
	(652)	(111)	(763)

Notes: All students take the same business, economics and law classes. They only have different engineering majors.

15% of the students are female. Foreigners are even more underrepresented with less than 4%. Table 2 provides descriptive statistics for the key measures.

Table 2: Descriptive statistics of baseline data set.

	Mean	Std. Dev.	Min.	Max.
Average University: $Grade_{i,j,k,t_1}$	2.8	.54	1.1	4.6
University entrance: $Grade_{i,j,k,\tilde{t}_{\leq -1}}$	1.9	.51	1	3.5
Female	.15	.35	0	1
Peer measure				
Next best$_{t_{-1},j}$ peer: $Peer_{i,j,k,\tilde{t}_{\leq -1}}$	2	.51	1	3.5
N	763			

Notes: 763 observations come from 763 students i, who first took their general university entrance examination (at a high school) at $\tilde{t}_{\leq -1}$, then took part in the group assignment with random peers at t_0. The students took part in the group assignment at 10 different dates between 2009 and 2016. They split up into 4 engineering majors j (see Table 1). 96.7% of the sample are Germans. The rest have one out of 11 other nationalities.

When students enter university, they are invited to an *orientation week* (OW), which takes place in the very first week of studies. With approx. 220 students in a cohort the considered field of study is very large and thereby anonymous. It is also important to point out that due to the good reputation of the university in the field, students are attracted from various regions of Germany. In many cases, they do not have an established social network of fellow students when entering the university. Students are randomly assigned to other students consisting of the same major with an average group size of eight students to spend one week together. During this week students are given an introduction to the university by senior students. The students' union organizes the event. The week usually culminates in a big party.

As an outcome measure we consider academic grades. It is important to understand the (unintuitive) German grade system with 1.0 (being the best grade) and the highest value 5.0 (being the worst grade).[10] Similar measures are applied both at high school and at university. The German high school degree which permits enrollment at university is called *Abitur* and its grade will be subsequently referred to as University Entrance Grade. The university grade is updated every semester. It considers all exams taken until this point and is computed as a weighted average following the curriculum rules. If an exam is failed, a grade of 5.0 is awarded.

Traditionally, the final grade is computed as the weighted average of individual exam results in the various courses and a final thesis conducted on one's own. Thus, team work exercise is not on the traditional agenda.

However, a mandatory group assignment labeled *Projekt im Bachelor* (engl. Project at Bachelor level, henceforth PB) is part of the curriculum. The PB intends to help students build soft skills and gain hands-on experience as group work is essential in the future professional life. During the PB, students have to come up with a business plan for some novel technological idea covering various aspects of marketing, budgeting, and legal implementation. After one week of intensive group work, all groups have to deliver a final report and pitch their results in front of a jury of professors and professionals. The task is deliberately designed with time pressure to induce cooperation between group members. Students are randomly assigned into groups.[11] Each group consisted of 11 to 13 students and each semester 14 to 17 groups participated.[12] We have access to PB data for the time period between 2009 and 2016 on 10 different dates.

All students from one project group either passed or failed the course. This design provides an opportunity for free-riding, which cannot be punished by the teaching staff. Only informal punishment of social pressure within the respective group can be exerted in order to control for free-riding. Spending an intensive week together with approx. 10 other students allows participants to get to know fellow students, especially their work ethos. Technically, the students are not forced to have any social interaction with other students from this time on. Nevertheless, it may be rewarding to be in contact

[10] 1.0 refers to excellent, 2.0 to good, 3.0 to satisfactory, 4.0 to passing, 5.0 to failure. In individual exams, further grading steps exist: 1.0, 1.3, 1.7, 2.0, 2.3, 2.7, 3.0, 3.3, 3.7, 4.0 and 5.0. As we use averages, all other intermediate steps (with up to two digits after the decimal point) are possible.

[11] After some negative experiences in the first year, the only exception to randomness is the distribution of females. Being in a minority, females found it hard to speak up in a group of males only. Since then, the organizers have stipulated that there are either two or more females in a group or no female at all.

[12] The only exception is our last cohort in 2015, for which a small cohort size allowed for group sizes of 8-9 individuals and 5 groups only.

with other students in order to efficiently prepare for exams and also to obtain informal information about classes.

This scenario differs relative to the OW scenario along several dimensions. First of all, OW takes place before the official course work starts. As such, students are more of a *blank slate* in terms of social connections. OW is deliberately designed as a social gathering. It is not organized by an official institution but by the students' union. As such, attendance is not mandatory. Students might not show up, thereby having little exposure to each other. Secondly, the non-mandatory nature might also induce a self-selection issue: e.g., low-ability students do not show up (although assigned to a group) and drop out later on. The PB is usually taken around the fifth semester, at which time it is very likely that students have already built up strong social networks.

It is important to compare our general institutional setting with the one prevailing in the United States, for which early influential studies exist (e.g. Zimmerman (2003) or Sacerdote (2001)).[13] In the United States, academic achievement is mostly proxied by the Scholastic Assessment Test (SAT) score from school (a standardized test covering math, writing, and reading) and the GPA score (grade point average) from university. More recently, Carrell et al. (2009) perform a similar exercise using data from Air Force Academy students. In this case, the relationship between the students is very close and students have little exposure to other outside contacts. In our setting, we use the purely random assignment of students to a class project.

In order to attend a German university, students have to have *Abitur* – the German equivalent to the SAT – or a foreign equivalent. To fill the given amount of places, the university picks the students with the best grade in Abitur. Given the prestigious nature of Technische Universität Darmstadt in general and the considered study field in particular, applications are received from all over Germany and also from abroad. As such students rarely know each other when entering the university. In general, studying in Germany is cost-free and there are no barriers related to the region of origin (either within Germany or along foreign vs. domestic students). The university of Darmstadt is not a campus university. This means above all that students do not usually live in university-provided housing. As a result, there is a lot of interaction of students with others outside their respective study class. In particular, there is also another higher-education institution in Darmstadt: the university of applied sciences, which, however,

[13]There are also several studies with Italian data (Brunello et al. (2010), Paola & Scoppa (2010), or De Giorgi et al. (2007)). As another example the study of Thiemann (2018) uses Swiss evidence.

has fewer students. The students in our sample might interact with students, from the latter institution in their free time, but not during their studies. In general, our setup provides less (forced) interaction than the usual US examples (especially the military academy case, cf. e.g. Lyle (2009)).

4. Empirical evaluation

In order to test our hypothesis, we run a regression of the following type:

$$G_{i,j,k,t_+} = a + b \cdot S_{i,t_-} + c \cdot Peer_{i,j,k,t_-} + FE_{j,t} + \epsilon_{i,j,k,t_+}, \tag{17}$$

for each individual i member of a peer group k in the major j (as presented in Table 1). We are most interested in the value of c, which captures both endogenous and exogenous peer effects jointly (as elaborated in Section 2.2). In our main investigation, we consider the mandatory PB project work. It is important to note the timing. In the baseline model, individuals get to know each other at the PB, which we denote as time t_0. The outcome is considered in a time after meeting t_+ and controlled for by effects preceding this event t_-, in which the connection did not exist. Our time resolution is one semester (half a year), implying that t_1 would refer to one semester after the initial encounter. The study focuses on outcomes during university enrollment. By construction, the University Entrance Grade is attained before all other measures and as such labeled by $\tilde{t}_{\leq -1}$. Furthermore, we consider fixed effects FE for the group assignment date t_0. These fixed effects capture heterogeneous effects for every single group assignment date that affect all participants of one cohort in the same way. The vector S_{i,t_-} includes individual explanatory measures prevailing before the establishment of the connection or not varying in time. In most specifications, we include the University Entrance Grade as well as dummies for gender and nationality. Usually, we consider the average university grade of the individual after being exposed to the peer group as the outcome G.

4.1. Baseline results

We consider different specifications for the peer effect $Peer$. The PB group defines the available connections. Given the different majors, it is more reasonable to connect to individuals of the respective major with a total overlap in study subjects, in particular since the courses from the respective engineering departments are considered tougher to study. As such, the peer group of an individual i in the PB group k only consists of the subset of individuals with the same major j. In the baseline, this definition gives rise to 209 peer groups. As elaborated in Section 2.2, the standard peer measure is the mean outcome in this peer group which takes place before the peer group was formed. Table A.1 in Appendix A summarizes the descriptive statistics of the underlying regression.

Table 3: OLS estimation of average university grade on university entrance peer grade.

				Randomness	Pseudo-peers
Average University:	$Grade_{i,j,k,t_1}$			$Grade_{i,j,k,t_{-3}}$	$Grade_{i,j,k,t_1}$
	(1)	(2)	(3)	(4)	(5)
Peer measures (based on Univ. entrance grade)					
Next best$_{t_{-1},j}$ peer: $Peer_{i,j,k,\tilde{t}_{\leq -1}}$	0.14**	0.17**	0.11*	0.07	
	(0.04)	(0.04)	(0.05)	(0.06)	
Next best$_{t_{-1},j}$ pseudo peer: $PseudoPeer_{i,j,k,\tilde{t}_{\leq -1}}$					0.08
					(0.06)
Control for individual i's ability					
University entrance: $Grade_{i,j,k,\tilde{t}_{\leq -1}}$	0.50**	0.52**	0.50**	0.61**	0.53**
	(0.04)	(0.04)	(0.05)	(0.06)	(0.05)
t_0 Group assignment date Fixed Effects	Yes	Yes	Yes	Yes	Yes
$t_0\times$ Major j Fixed Effects	No	Yes	Yes	Yes	Yes
Adj. R^2	0.28	0.29	0.29	0.26	0.28
N	763	763	504	504	321
# Peer groups	209	209	187	187	145

Robust *SE* in parentheses; * $p < 0.05$, ** $p < 0.01$

Notes: In column (1)-(3) and (5), the dependent variable is the average university grade of student i in major j at time t_1 (which is after group assignment). In column (4), the dependent variable is the same but at time t_{-3} (before group assignment). The peer measures are based on the university entrance grade at $\tilde{t}_{\leq -1}$ (which was before going to University [i.e., even before t_{-3}]). k determines the group into which student i is randomly allocated for the group assignment. We include (but do not show) a gender dummy and country of origin dummies in all regressions. Columns (2) and (3) only differ in sample size. In column (3), we use the same sample as in column (4). In column (4), we loose observations because the dependent variable $\left(Grade_{i,j,k,t_{-3}}\right)$ is not defined at t_{-3} if a student take part in the mandatory PB group event before his or her fourth semester of studies.

We present this regression in Table A.2 in column (1) of Appendix A.1, showing no significant effect. This is in line with the reasoning presented in Section 2. Given the lack of enforced interaction, individuals can avoid their peers from the PB group and self-select their peers. As formally presented in Section 2, there is an incentive to self-select an individual who is of similar ability.

As such, there will be a pair formation along pre-observed ability. We use this idea to construct a *next-best* peer measure. From the given set of individuals in the peer group k with the same major j individuals are assigned a peer who is closest to them in terms of academic achievement before getting to know each other $\tilde{t}_{\leq -1}$.[14] The peer measure itself is constructed with the University Entrance Grade (time index $\tilde{t}_{\leq -1}$). As shown in Table 3 column (1), this produces a significant peer effect in line with our theoretical reasoning.[15] This result is robust to including a fixed effect that interacts the cohort and the major (cf. column (2) of Table 3) to control for cohort-major specific effects. As such, the next-best peer effect is also robust to controlling for heterogeneous ability with respect to majors by cohort. From a quantitative perspective this finding suggests that the increase in the peer quality by one standard deviation (-0.5 grade steps) increases the performance of the individual by -0.17 standard deviations (equivalent to a grade step of approx. -0.1). Note that in the German grade system 1.0 is the best grade and 5.0 the worst grade.

In order to validate these results, we have to check whether the assignment to the groups and, even more so, the construction of the peers is truly random. As such, we run a regression of a similar type to the regression in equation 17 with the only difference that both the right hand side (the explained variable) and the explanatory factors are measured before t_0, which is the time in which individuals work together during PB. As individuals were not (formally) introduced by the PB, they should not exert an influence on each other at this point in time. In order to run this regression, we should focus on exactly the same sample in the baseline regression and in the randomness test.[16] As reported in the Table 3 column (3) the overall result of a significant peer effect remains stable for this sample. The randomness check (column (4) of Table 3), however, shows that no peer effect exists before individuals worked together in a PB group.

As an alternative to test the randomness of the assignment we form a *pseudo peer measure*. In this case, we randomize the group formation ourselves and then construct the next-best peer measure to assign individuals a Pseudo peer from a random group. As displayed in column (5) of Table 3 this Placebo regression also does not show a significant

[14] Formally, this is the individual for whom the absolute deviation in average grades before meeting is the lowest. As a result, this will not construct exact pairs. While the first best is connected to the second best, the second best herself might be closer (according to the given definition) to the third best and so on.

[15] It is important to point out that we only know potential peers given the common exposure. We cannot, however, guarantee that individuals actually studied together in groups. A way of testing this would be to conduct a survey asking students about their study partners. Unfortunately, this is not possible as the students covered in the sample have left university by now.

[16] We loose some observations in column (4) of Table 3 because some students take part in the mandatory PB group event before their fourth semester of studies. In this case, the dependent variable $(Grade_{i,j,k,t_{-3}})$ is not defined (at t_{-3}).

peer effect, which again confirms the randomness of the initial assignment.[17] Of course, a single non-significant regression can be a chance result. To overcome this issue, we run 1,000 random pseudo peer group generations and then our standard regression for which the mean relevant coefficient across repetitions amounts to 0.0286 (median: 0.0567). Figure A.1 in Appendix A.2 reports the associated t-statistics to consider significance. For the considered 1,000 repetitions only 61 cases (60 times $t \geq 2$ and once $t \leq -2$) exceed the absolute value of 2 associated with a confidence level of approximately 5% with a mean t-statistic of 0.503 (median: 0.47). Thus, the results are also robust to the Placebo treatment.

4.1.1. Robustness checks

To further validate our baseline result reported in Table 3 column (2), we conduct several robustness checks. Our measure of achievement is the average university grade. By construction, this measure may be auto-correlated in time as it aggregates all exams taken (both passed and failed) until a certain time point and computes an average. Table A.3 (Appendix A.3) discusses the robustness of our findings to alternative performance specifications. A simple way to isolate outcomes which take place after the peer group was formed from initial outcomes would be to control for a lagged weighted average of the university grade. As shown in column (1) of Table A.3 (Appendix A.3) with a lagged time of three semesters, the lagged value is picked up as significant. The overall peer effect also remains significant, albeit a bit smaller in magnitude. Just controlling for a lagged university grade may still be problematic, as in the earlier years of studies average grades vary a lot (due to the small number of exam grades considered in the average university grade), whereas they are less volatile once students proceed further with their studies. To control for this, we include study progress times major fixed effects. These capture heterogeneous levels of the average university grade at different study progress times by major and make average university grades comparable in these dimensions.[18] The results produced under this specification (column (2) in Table A.3, Appendix A.3) are very close to the baseline specification, confirming that the outcomes are not the result of a measurement error.

[17] The number of observations is lower with Pseudo-peers since random peers may be from different cohorts and we still require the correct order of events.

[18] With $t_0 \times$ Major$j$ fixed effects, we control for heterogeneity across 25 group. With $t_0\times$ Major j $\times$ Study progress fixed effects, we control for heterogeneity across 90 groups.

As a final robustness check in this dimension, we can replace the average university grade as our final outcome measure by a specific exam. We use the mandatory exam in Microeconomics I for those individuals who take this exam after working in a PB group together. We consider this exam representative for the whole curriculum as it is one of the non-technical courses mandatory for all majors, but also tests formal reasoning and mathematical skills. The exam results are measured in points (for which a high number of points suggests a better result). Given that in the German grade system a high value is a sign of low ability, we find a negative correlation (see column (3) in Table A.3, Appendix A.3). The correlation remains significant, suggesting that the peer effect is also robust to this alternative specification. We can also use the order of events for a natural randomness test. Some students wrote Microeconomics I first and then did PB. For these, we identify no peer effect. This randomness test (column (4) of Table A.3, Appendix A.3) once again suggests that the effect is not prevailing before individuals work together in a PB group.[19]

4.1.2. Different peer measures

Dropping the assumption that peer groups are only formed along identical majors still maintains a significant peer effect (cf. Table A.2 column (2) in Appendix A.1), even slightly increasing the magnitude of the measured effect. As a result, we can say that the peer effect is not bound to the specific major.

While the standard mean peer effect implicitly assumes a social network topology, in which everyone is connected to everyone, our baseline specification indicates a line type topology, in which individuals are connected in a line. Thus, we can also define a *next second best*, i.e. an individual to whom an individual is not directly connected, but indirectly through her direct peer standing next in the line. As shown in Table A.2 column (3) in Appendix A.1, this measure of peer effects, however, does not produce significant results implying that the effect quickly abates along the line.

A different network structure that is frequently discussed in the literature is that of the star type with one individual who is connected to every other individual. This key player in the group exerts an effect on the rest of the individuals. If the group is dominated by the best individual, this is often labeled a *shining light* effect. In the polar case, in which the worst individual dominates the group, this is frequently referred to as

[19]The results in columns (4) and (5) of Table A.3 (Appendix A.3) are robust to focusing on exactly the same sample.

the *bad apple* effect. The latter is highly relevant for school education where students learn together in small classes (e.g. Kristoffersen et al. (2015) or Carrell et al. (2018)) and it is thus hard to overcome the negative influence.

We identify the potential *shining light* of the group as the individual with the best result based on the University Entrance Grade. As shown in Appendix A.4 in Table A.4 column (1), in this setting we cannot observe a significant peer effect.[20]

We can contrast this with a *bad apple* effect. We identify the latter as the individual with the worst University Entrance Grade. In this scenario, we find a significant positive effect (column (2) of Table A.4, Appendix A.4). This result passes the usual randomness check (column (4)) and also holds in the sample employed for the randomness check (column (3)). If we relax the assumption of individuals also being within the same major j (column (5) of Table A.4, Appendix A.4), this result still ceases to hold. The *bad apple* effect is less robust than the overall finding. Yet, it is stronger than the *shining light* effect, showing that the worst individuals exert a stronger impact on the rest of the group than the best within the group.

4.2. Group work vs. orientation week

We want to contrast the scenario of the mandatory PB group work with a different intervention. Before the official courses start, students are invited to partake in an *orientation week*, allowing them to get to know potential study partners.

The properties of the considered sample are reported in Table B.5 and B.6 in Appendix B, revealing similar properties as in the PB sample (cf. Tables 1 and 2). The sample size for the PB case is a bit smaller, as by the time it takes place a substantial amount of individuals have already dropped out of the study program. We slightly modify the notation and define our reference date as $\tilde{t}_0$: the OW takes place after high school ends and shortly before university begins.

Table 4 presents the results of the baseline regression for the new sample. Neither the *next-best* peer effect (column (1) of Table 4) nor the standard mean peer effect (column (5)) are significant. This is in line with the theoretical reasoning. In the social setting individuals might not find learn about academic abilities of their peers and thus

[20]The result even weakens if we relax the assumption that the peer must be of the same major (column (5) in Table A.4, Appendix A.4.

not be able to find matching learning groups. As a result, these learning groups likely cease to exist after a while not creating a measurable peer effects.

The results also does not change by altering the structure of the fixed effects in the way presented earlier (column (2) of Table 4). One important point already acknowledged is that the sample consists of both students who successfully finish the program and those who quit the program. As a further check, we decompose these two groups, leaving us with a roughly equal-size split of the sample (cf. columns (3) and (4) of Table 4). Once again no significant results emerge.

Table 4: OLS estimation of average university grade on university entrance peer grade. Peer group formation is during the orientation week (the very first week of studies).

	All		Successful	Unsuccessful	Mean
	$Grade_{i,j,k,\tilde{t}_1}$				
	(1)	(2)	(3)	(4)	(5)
Peer measures (based on Univ. entrance grade)					
Next best$_{\tilde{t}_{\leq -1},j}$ peer: $Peer_{i,j,k,\tilde{t}_{\leq -1}}$	-0.13	-0.14	-0.19	-0.10	
	(0.14)	(0.14)	(0.17)	(0.26)	
Mean peers: $Peer^{mean_j}_{k,\tilde{t}_{\leq -1}}$					-0.19
					(0.15)
Control for individual i's ability					
University entrance: $Grade_{i,j,k,\tilde{t}_{\leq -1}}$	0.77**	0.77**	0.75**	0.81**	0.67**
	(0.13)	(0.13)	(0.16)	(0.23)	(0.06)
$\tilde{t}_0$ Group assignment date Fixed Effects	Yes	Yes	Yes	Yes	Yes
$\tilde{t}_0\times$ Major j Fixed Effects	Yes	No	Yes	Yes	Yes
Adj. R^2	0.18	0.16	0.19	0.20	0.18
N	932	932	494	438	932
# Peer groups pre-group assignment	119	119	76	66	119

Robust *SE* in parentheses; * $p < 0.05$, ** $p < 0.01$

Notes: In column (1)-(5), the dependent variable is the average university grade of student i in major j at time $\tilde{t}_1$ (which is after group assignment). The peer measures are based on the university entrance grade at $\tilde{t}_{\leq -1}$ (which was before going to University [i.e., before $\tilde{t}_0$]). k determines the group into which student i is randomly allocated for the group assignment. We include (but do not show) a gender dummy and country of origin dummies in all regressions.

We also run (but do not show) regressions with the other alternative peer measures used in the PB context. Again, we do not find any indication for peer effects in the OW setting.

We can thus conclude that this intervention – in contrast to the former (cf. Section 4.1) – does not produce peer effects. Thus, professional bonds influencing future academic success are formed in professional settings (a formal mandatory group project) rather than during informal social events.

4.3. Long-run peer effects

The group events, in which students get to know each other, only lasts for one week. So far we documented that this, however, will have an effect on individual study outcomes one semester later (in case of the mandatory PB group work). There is no effect for the informal social event (OW). Thiemann (2018) considers a related setup, in which individuals are randomly assigned to a group at the beginning of their studies and documents that these effects are, however, persistent.

We conduct a similar exercise in Table 5. Basically, we start with the baseline regression (cf. eq. 17) but look at outcomes that are farther away from the date of PB group work. By construction, the sample size decreases if we focus on long-term effects: individuals drop out of the sample because of quitting the study program, either successfully or unsuccessfully. For all cases, three semesters (column (1)), five semesters (column (2)), and even seven semesters afterwards (column (3)), the next-best peer effect is still present and significant and only slightly abates in time. Hence, we can confirm the long-run effect of the short-term PB intervention.[21]

In line with our previous findings for the social group event, we do not find a next-best peer effect farther away from the OW. This result holds for three semesters (column (4) in Table 5), five semesters (column (5)), and even seven semesters (not shown) after OW.

[21] In contrast, Thiemann (2018) finds that the average outcome of the peer group has a negative effect on the individual suggesting (long-run) discouragement effects.

Table 5: Persistence over time for OLS estimation of average university grade on university entrance peer grade.

	Mandatory group event (PB)			Social group event (OW)	
	Mid-term	Long-term	Long-term II	Mid-term	Long-term
	$Grade_{i,j,k,t_3}$	$Grade_{i,j,k,t_5}$	$Grade_{i,j,k,t_7}$	$Grade_{i,j,k,\tilde{t}_3}$	$Grade_{i,j,k,\tilde{t}_5}$
	(1)	(2)	(3)	(4)	(5)
Peer measures (based on Univ. entrance grade)					
Next best$_{t_{\leq -1},j}$ peer: $Peer_{i,j,k,\tilde{t}_{\leq -1}}$	0.15**	0.16**	0.15**		
	(0.04)	(0.05)	(0.05)		
Next best$_{\tilde{t}_{\leq -1},j}$ peer: $Peer_{i,j,k,\tilde{t}_{\leq -1}}$				-0.10	-0.13
				(0.10)	(0.08)
Control for individual i's ability					
University entrance: $Grade_{i,j,k,\tilde{t}_{\leq -1}}$	0.46**	0.48**	0.51**	0.58**	0.62**
	(0.04)	(0.04)	(0.04)	(0.08)	(0.07)
t_0 Group assignment date Fixed Effects	Yes	Yes	Yes		
$t_0\times$ Major j Fixed Effects	Yes	Yes	Yes		
$\tilde{t}_0$ Group assignment date Fixed Effects				Yes	Yes
$\tilde{t}_0\times$ Major j Fixed Effects				Yes	Yes
Adj. R^2	0.23	0.25	0.28	0.23	0.23
N	746	717	694	931	929
# Peer groups	209	208	200		
# Peer groups pre-group assignment				119	119

Robust *SE* in parentheses; * $p < 0.05$, ** $p < 0.01$

Notes: In columns (1), (2) and (3), the dependent variable is the average university grade of student i in major j at time t_3, t_5 or t_7 (which is always after PB group assignment). In columns (4) and (5), the dependent variable is the average university grade at time $\tilde{t}_3$ or $\tilde{t}_5$ (which is always after OW group assignment). The peer measures are based on the university entrance grade at $\tilde{t}_{\leq -1}$ (which was before going to University [i.e., even before t_{-1} or t_{-3}]). k determines the group into which student i is randomly allocated for the group assignment. We include (but do not show) a gender dummy and country of origin dummies in all regressions.

4.4. Heterogeneous peer effects

After the initial observation on the existence of peer effects, researchers have tried to gain a more differentiated picture of whether certain subgroups profit more (or less) from peer effects. For example, Griffith & Rask (2014) argue that male and minority students are more affected by their peers.

In our data set, females are a minority (cf. Table 2). This is not specific to Technische Universität Darmstadt, but more generally to studies in Engineering, and Natural Science in Germany. There are several state-financed initiatives in order to improve the female share in these subjects. One problem might lie in the fact that female students find it hard to assert themselves in a male-dominated peer environment.

In order to test this claim in our setting, we split the sample into male and female students just to run our standard regression (cf. equation 17). While the explained variables are split by gender, it might still be (especially for the minority of females) that the identified next-best peer is of a different gender. While the peer effect is robust and of similar magnitude as before for the male individuals (cf. column (1) of Table 6), it ceases to exist for female students (column (2)). This result supports the idea that females find it harder to integrate or to be integrated in a male-dominated environment. This finding is also in line with the results of Griffith & Rask (2014), which show that males are more influenced by peers. However, in their setting they consider a more gender-balanced environment. As such, the lack of peer influence for females must not be a sign of discrimination, but can also be the result of gender-specific learning types. Maybe females are more likely to work in solitude or form peers in a way we cannot identify.

Table 6: OLS estimation with peer measures by gender.

	Mandatory group event (PB)		Social group event (OW)	
	Male	Female	Male	Female
	$Grade_{i,j,k,t_1}$		$Grade_{i,j,k,\tilde{t}_1}$	
	(1)	(2)	(3)	(4)
Peer measures (based on Univ. entrance grade)				
Next best$_{t_{\leq -1},j}$ peer: $Peer_{i,j,k,\tilde{t}_{\leq -1}}$	0.16**	0.04		
	(0.04)	(0.10)		
Next best$_{\tilde{t}_{\leq -1},j}$ peer: $Peer_{i,j,k,\tilde{t}_{\leq -1}}$			-0.11	-0.64
			(0.12)	(0.33)
Control for individual i's ability				
University entrance: $Grade_{i,j,k,\tilde{t}_{\leq -1}}$	0.51**	0.60**	0.68**	1.17**
	(0.04)	(0.12)	(0.10)	(0.32)
t_0 Group assignment date Fixed Effects	Yes	Yes		
$t_0 \times$ Major j Fixed Effects	Yes	Yes		
$\tilde{t}_0$ Group assignment date Fixed Effects			Yes	Yes
$\tilde{t}_0 \times$ Major j Fixed Effects			Yes	Yes
Adj. R^2	0.30	0.35	0.23	0.22
N	652	111	730	202
# Peer groups	203	98		
# Peer groups pre-group assignment			119	88

Robust SE in parentheses; * $p < 0.05$, ** $p < 0.01$

Notes: Across columns, the dependent variable is the average university grade of student i in major j at time t_1 (or for columns (3) and (4) at $\tilde{t}_1$) which is after group assignment. The peer measures are based on the university entrance grade at $\tilde{t}_{\leq -1}$ (which was before going to University [i.e., even before t_{-1} or t_{-3}]). k determines the group into which student i is randomly allocated for the group assignment. We include (but do not show) a gender dummy and country of origin dummies in all regressions.

We also run the same specification for our second intervention – the orientation week, which is an informal social group event – but once again cannot find significant peer effects even after disentangling for gender (cf. columns (3) and (4), Table 6).

We also consider heterogeneous peer effects with respect to ex-ante ability. In order to do so, we split individuals into those above median in term of University Entrance Grade and below. It turns out that those individuals who are already good to begin with also gain more from their peers (cf. columns (1) and (2) of Table 7) when considering the PB event. Once again, we find no significant effect for the social group event (see columns (3) and (4)).[22]

Table 7: OLS estimation with peer measures by ex-ante ability.

	Mandatory group event (PB)		Social group event (OW)	
Relation to Median	Below	Above	Below	Above
	$Grade_{i,j,k,t_1}$		$Grade_{i,j,k,\tilde{t}_1}$	
	(1)	(2)	(3)	(4)
Peer measures (based on Univ. entrance grade)				
Next best$_{t_{\leq -1},j}$ peer: $Peer_{i,j,k,\tilde{t}_{\leq -1}}$	0.12*	0.26**		
	(0.05)	(0.06)		
Next best$_{\tilde{t}_{\leq -1},j}$ peer: $Peer_{i,j,k,\tilde{t}_{\leq -1}}$			-0.10	-0.30
			(0.16)	(0.16)
Control for individual i's ability				
University entrance: $Grade_{i,j,k,\tilde{t}_{\leq -1}}$	0.32**	0.91**	0.40**	1.19**
	(0.05)	(0.12)	(0.15)	(0.17)
t_0 Group assignment date Fixed Effects	Yes	Yes		
$t_0\times$ Major j Fixed Effects	Yes	Yes		
$\tilde{t}_0$ Group assignment date Fixed Effects			Yes	Yes
$\tilde{t}_0\times$ Major j Fixed Effects			Yes	Yes
Adj. R^2	0.13	0.24	0.13	0.15
N	417	346	524	408
# Peer groups	195	152		
# Peer groups pre-group assignment			115	115

Robust *SE* in parentheses; * $p < 0.05$, ** $p < 0.01$

Notes: Across columns, the dependent variable is the average university grade of student i in major j at time t_1 (or for columns (3) and (4) at $\tilde{t}_1$) which is after group assignment. The peer measures are based on the university entrance grade at $\tilde{t}_{\leq -1}$ (which was before going to University [i.e., even before t_{-1} or t_{-3}]). k determines the group into which student i is randomly allocated for the group assignment. We include (but do not show) a gender dummy and country of origin dummies in all regressions.

This finding is in line with the reasoning put forward in Section 2.2, arguing that individuals with above-average grades profit more from a one-on-one matching as compared to a classroom setting. Thus, one can conclude this section by stating that peer effects mostly benefit individuals who are already successful and in a majority (males). In consequence, the observed pair-wise peer effects do not promote inclusion or equity in outcomes.

[22] Note that we only have University Entrance Grades with one digit after the decimal point. Therefore, we have several observations with the median value and the sub-samples below, and above median are not of the same size.

4.5. Discussion

As peer interaction has a positive impact on individuals, it is likely that policy makers (e.g. directors of study programs) would like to impose it. We consider the two settings of the officially imposed group work and the more informal orientation week at the very beginning of studies, from which only the formal implies the presence of a peer effect. The lower impact of the orientation week is likely to be found in its informal nature. We have information on group formation taking place on the first date of the week. As there is no formal mechanism for punishing non-attendance, only a subset of the students interact during this week. Secondly, the timing of the event matters. Being literally in their first week, students are expected not to have formed any alternative social connections. On the other hand, as revealed in the sub-samples displayed in Table 4 there is a high dropout rate: only about half of the sample successfully finish the study program. Thus, many students you get to know in the first week are very likely not to be around for long.

In contrast, the more official and later intervention in the form of the mandatory group project displays a substantial effect. Even in this setting the peer effect is rather narrow, only running between individuals that are similar ex-ante. Moreover, it only persists for male students. This could be considered a sign that female students – being a minority in our setting – systematically form other peer networks. As this is, however, also in line with Griffith & Rask (2014), who investigate more gender-balanced student groups, this could also signify that males *per se* are more prone to social influences.

The lack of substantial peer effects might be considered a failure; in particular, if the interventions were designed to push underachieving students. As shown in Section 2.2, homophilic pairwise matching is inferior to overall group effects (mean peer effects) for individuals who are below the (group) average. Yet, as was also shown in Section 2.2, under the presence of peers all individuals are better off than in autarky.

In our empirical investigation the result is likely to originate from the overall anonymous structure of the German university system that – for our specific setup – does not enforce interaction besides the cases considered in this paper. In general, the German university system emphasizes the autonomous and self-managed nature of studies. The more formalized structure – including the mandatory group work – was introduced in the Bologna Process in order to improve the comparability of the German higher education system to the one in other European countries.[23] In general, the latter induced features that used to be more associated with high schools in Germany. Nevertheless, the trend is to introduce more online learning contents, which even makes the physical presence at the lecture – and thereby exposure to other students – unnecessary. As such, peer exposure and the resulting effects may decrease.

5. Conclusion

In this paper, we document a peer effect in higher education within the highly anonymous German university system. We identify the peers from a mandatory group assignment in which individuals were randomly assigned to groups. We exploit pre-university grades since peers did not know and therefore cannot have influenced each other at this stage. The group assignment proves to have long-run effects. Rather than having an overall impact, however, we document that the effect is only limited to individuals who are similar ex-ante. As such, the role of social engineering is limited in this setting (Carrell et al., 2013).

Our results are coherent with a theoretical reasoning showing that it is rational to match with individuals who are already of similar ability. We only find a robust next-best peer effect in line with a network structure of the line type. Other network structures such as a star network and especially a fully connected network do not suggest the existence of peer effects.

We contrast our finding for the mandatory group work with another informal social group event, which also leads to random group formation. The social group event takes place in the very first week of studies and is organized by the students' union. Here, we cannot measure any significant peer effects.

Our setup could be extended in various ways. A larger share of students end up in managing positions of the largest German companies. Shue (2013) considers the network of Harvard Business School students and shows that it extends over into the professional realm several years after graduating. It would be interesting to cover the post-graduate period, potentially yielding important insights with a view to structuring alumni networks managed by a university.

[23]Initiated in 1999, the German degree programs were replaced by Bachelor and Master Programs in line with the Bologna Process.

References

Back, M. D., S. C. Schmukle, & B. Egloff (2008). Becoming friends by chance. *Psychological Science 19*(5), 439–440. PMID: 18466403. doi:10.1111/j.1467-9280.2008.02106.x.

Bénabou, R. (1996). Heterogeneity, stratification, and growth: Macroeconomic implications of community structure and school finance. *The American Economic Review 86*(3), 584–609. Available from: http://www.jstor.org/stable/2118213.

Blume, L. E., W. A. Brock, S. N. Durlauf, & Y. M. Ioannides (2011). Chapter 18 - Identification of social interactions. Volume 1 of *Handbook of Social Economics*, pp. 853–964. North-Holland. doi:10.1016/B978-0-444-53707-2.00001-3.

Brady, R., M. Insler, & A. Rahman (2016). Bad Company: Reconciling Negative Peer Effects in College Achievement. MPRA Paper 68354, University Library of Munich, Germany. Available from: https://ideas.repec.org/p/pra/mprapa/68354.html.

Brunello, G., M. De Paola, & V. Scoppa (2010). Peer effects in higher eduaction: Does the field of study matter? *Economic Inquiry 48*(3), 621–634. doi:10.1111/j.1465-7295.2009.00235.x.

Burke, M. A. & T. R. Sass (2013). Classroom peer effects and student achievement. *Journal of Labor Economics 31*(1), 51–82. doi:10.1086/666653.

Carrell, S. E., R. L. Fullerton, & J. E. West (2009). Does Your Cohort Matter? Measuring Peer Effects in College Achievement. *Journal of Labor Economics 27*(3), 439–464. doi:10.1086/600143.

Carrell, S. E., M. Hoekstra, & E. Kuka (2018). The long-run effects of disruptive peers. *American Economic Review 108*(11), 3377–3415. doi:10.1257/aer.20160763.

Carrell, S. E., B. I. Sacerdote, & J. E. West (2013). From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation. *Econometrica 81*(3), 855–882. doi:10.3982/ECTA10168.

De Giorgi, G., M. Pellizzari, & S. Redaelli (2007). Be as Careful of the Books You Read as of the Company You Keep: Evidence on Peer Effects in Educational Choices.

IZA Discussion Papers 2833, Institute for the Study of Labor (IZA). Available from: `https://ideas.repec.org/p/iza/izadps/dp2833.html`.

Epple, D. & R. E. Romano (2011). Chapter 20 - Peer effects in education: A survey of the theory and evidence. Volume 1 of *Handbook of Social Economics*, pp. 1053–1163. North-Holland. `doi:10.1016/B978-0-444-53707-2.00003-7`.

Griffith, A. L. & K. N. Rask (2014). Peer effects in higher education: A look at heterogeneous impacts. *Economics of Education Review 39*(C), 65–77. `doi:10.1016/j.econedurev.2014.01.003`.

Kang, C. (2007). Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea. *Journal of Urban Economics 61*(3), 458–495. `doi:10.1016/j.jue.2006.07.006`.

Kristoffersen, J. H. G., M. V. Krægpøth, H. S. Nielsen, & M. Simonsen (2015). Disruptive school peers and student outcomes. *Economics of Education Review 45*, 1–13. `doi:10.1016/j.econedurev.2015.01.004`.

Lyle, D. S. (2009). The effects of peer group heterogeneity on the production of human capital at West Point. *American Economic Journal: Applied Economics 1*(4), 69–84. Available from: `http://www.jstor.org/stable/25760182`.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies 60*(3), 531–542. Available from: `http://www.jstor.org/stable/2298123`.

Oosterbeek, H., A. S. Booij, & E. Leuven (2016). Ability Peer Effects in University: Evidence from a Randomized Experiment. *The Review of Economic Studies 84*(2), 547–578. `doi:10.1093/restud/rdw045`.

Paola, M. D. & V. Scoppa (2010). Peer group effects on the academic performance of Italian students. *Applied Economics 42*(17), 2203–2215. `doi:10.1080/00036840701765478`.

Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *The Quarterly Journal of Economics 116*(2), 681–704. `doi:10.1162/003355301151144131`.

Sacerdote, B. (2011). *Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?*, Volume 3 of *Handbook of the Economics of Education*, Chapter 4, pp. 249–277. Elsevier. doi:10.1016/B978-0-444-53429-3.00004-1.

Shue, K. (2013). Executive Networks and Firm Policies: Evidence from the Random Assignment of MBA Peers. *Review of Financial Studies 26*(6), 1401–1442. doi:10.1093/rfs/hht019.

Thiemann, P. (2018). The Persistent Effects of Short-Term Peer Groups in Higher Education. Working Papers 2018:32, Lund University, Department of Economics. Available from: https://ideas.repec.org/p/hhs/lunewp/2018_032.html.

WirtschaftsWoche (2014). Hochschulranking 2014 - Das sind Deutschlands beste Unis. Website (last visit 6.4.2017). Available from: http://www.wiwo.de/erfolg/campus-mba/hochschulranking-2014-das-sind-deutschlands-beste-unis/10076006.html.

WirtschaftsWoche (2016). Hochschulranking 2016 - Das sind Deutschlands beste Unis. Website (last visit 6.4.2017). Available from: http://www.wiwo.de/erfolg/campus-mba/hochschulranking-2016-praxisvortraege-sind-das-salz-in-der-suppe/14719390-3.html.

Zimmerman, D. J. (2003). Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *The Review of Economics and Statistics 85*(1), 9–23. doi:10.1162/003465303762687677.

ONLINE-APPENDIX

The following should be supplementary material, only available online.

A. Details and robustness checks for PB group work

Table A.1 summarizes the descriptive statistics of the non-baseline regressions for the mandatory group work.

Table A.1: Descriptive statistics for robustness analysis.

	Mean	Std. Dev.	Min.	Max.	N
Average University: $Grade_{i,j,k,t_{-3}}$	3	.69	1	5	521
Average University: $Grade_{i,j,k,t_3}$	2.6	.55	1	4.3	746
Average University: $Grade_{i,j,k,t_5}$	2.5	.58	1	4.5	717
Average University: $Grade_{i,j,k,t_7}$	2.3	.61	1	5	694
Points: $Exam_{i,j,k,t_{-3}-t_7}$	58	15	8.5	89	725
Peer measures (based on Univ. entrance grade)					
Next best$_{t_{-1},j}$ pseudo peer: $PseudoPeer_{i,j,k,\tilde{t}_{\leq -1}}$	1.9	.51	1	3.4	321
Next best$_{t_{-1}}$ peer: $Peer_{i,j,k,\tilde{t}_{\leq -1}}$	2	.5	1	3.5	763
Next second best$_{t_{-1},j}$ peer: $Peer_{i,j,k,\tilde{t}_{\leq -1}}$	1.9	.46	1	3.4	628
Mean peers: $Peer^{mean_j}_{k,\tilde{t}_{\leq -1}}$	2	.37	1	3.5	763
Mean peers: $Peer^{mean}_{k,\tilde{t}_{\leq -1}}$	2	.28	1.3	3.1	763
Best$_{t_{-1},j}$ peer: $Peer_{j,k,\tilde{t}_{\leq -1}}$	1.7	.48	1	3.4	511
Worst$_{t_{-1},j}$ peer: $Peer_{j,k,\tilde{t}_{\leq -1}}$	2.1	.46	1.1	3.5	520
N	763				

Notes: 763 observations come from 763 students i, who first took their general university entrance examination (at a high school) at $\tilde{t}_{\leq -1}$, then took part in the group assignment with random peers at t_0. The students took part in the group assignment at 10 different dates between 2009 and 2016. They split up into 4 engineering majors j (see Table 1).

A.1. Other peer measures

The standard peer measure often used in the literature is the mean outcome in a peer group which takes place before the peer group was formed. We present a regression with this measure in Table A.2 in column (1), showing no significant effect.

In contrast to our baseline results (cf. Table 3 column (1)), we drop the assumption that peer groups are only formed along identical majors in Table A.2 column (2). In this specification, we still find a significant peer effect, even slightly increasing the magnitude of the measured effect.

We can also define a *next second best*, i.e. an individual to whom an individual is not directly connected, but indirectly through her direct peer standing next in the line. As shown in column (3), this measure of peer effects, however, does not produce significant results implying that the effect quickly abates along the line.

Table A.2: OLS estimation with alternative peer measures and robustness tests.

	Mean	All majors	2nd best
Average University:		$Grade_{i,j,k,t_1}$	
	(1)	(2)	(3)
Peer measures (based on Univ. entrance grade)			
Mean peers: $Peer^{mean_j}_{k,\tilde{t}_{\leq -1}}$	-0.05		
	(0.06)		
Next best$_{t_{-1}}$ peer: $Peer_{i,j,k,\tilde{t}_{\leq -1}}$		0.25**	
		(0.04)	
Next second best$_{t_{-1},j}$ peer: $Peer_{i,j,k,\tilde{t}_{\leq -1}}$			-0.01
			(0.04)
Control for individual i's ability			
University entrance: $Grade_{i,j,k,\tilde{t}_{\leq -1}}$	0.55**	0.50**	0.56**
	(0.04)	(0.04)	(0.04)
t_0 Group assignment date Fixed Effects	Yes	Yes	Yes
$t_0\times$ Major j Fixed Effects	Yes	Yes	Yes
Adj. R^2	0.27	0.31	0.29
N	763	763	628
# Peer groups	209	136	140

Robust *SE* in parentheses; * $p < 0.05$, ** $p < 0.01$

Notes: Across columns, the dependent variable is the average university grade of student i in major j at time t_1 (which is after group assignment). The peer measures are based on the university entrance grade at $\tilde{t}_{\leq -1}$ (which was before going to University [i.e., even before t_{-1} or t_{-3}]). k determines the group into which student i is randomly allocated for the group assignment. We include (but do not show) a gender dummy and country of origin dummies in all regressions.

A.2. Pseudo peer groups

Figure A.1 reports the t-statistics for 1,000 random pseudo peer group generations and corresponding runs of our standard regression. The mean relevant coefficient across repetitions amounts to 0.0286 (median: 0.0567). The results are robust to this Placebo treatment.

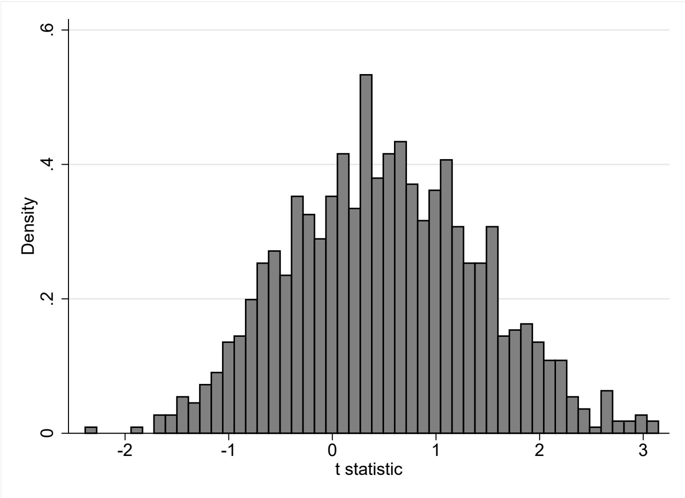

Figure A.1: Histogram of t-statistics for peer coefficient with pseudo peer exposure for 1,000 repetitions.

For the Placebo test, the mean of the t-statistic across 1,000 repetitions is 0.503, which indicates insignificance to relevant levels. Only in 60 cases the t-statistic exceeds the threshold value of 2 and in one case it undercuts the value of -2 (associated with a statistical significance at the 5% level).

A.3. Alternative performance measures

Table A.3 discusses the robustness of our findings to alternative performance specifications. As shown in column (1) of Table A.3 with a lagged time of three semesters, the lagged value is picked up as significant. The overall peer effect also remains significant, albeit a bit smaller in magnitude. The results produced with study progress times major fixed effects (column (2)) are very close to the baseline specification, confirming that the outcomes are not the result of a measurement error. We can also replace the average university grade as our final outcome measure by a specific exam. Given that in the German grade system a high value is a sign of low ability, we find a negative correlation (see column (3)). The correlation remains significant, suggesting that the peer effect is also robust to this alternative specification. Column (4) uses the order of events for a natural randomness test. Some students wrote Microeconomics I first and then did PB. For these, we identify no peer effect. This randomness test once again suggests that the effect is not prevailing before individuals work together in a PB group.

Table A.3: Robustness of OLS estimation of average university grade on university entrance peer grade – Alternative performance measures.

	Control grade	Study progress	Exam	Randomn. exam
	$Grade_{i,j,k,t_1}$		$Exam_{i,j,k,t_1-t_7}$	$Exam_{i,j,k,t_{-1}-t_{-3}}$
	(1)	(2)	(3)	(4)
Peer measures (based on Univ. entrance grade)				
Next best$_{t_{-1},j}$ peer: $Peer_{i,j,k,\tilde{t}_{\leq -1}}$	0.06*	0.17**	-5.22**	0.61
	(0.03)	(0.04)	(1.64)	(1.46)
Control for individual i's ability				
University entrance: $Grade_{i,j,k,\tilde{t}_{\leq -1}}$	0.14**	0.52**	-10.18**	-10.97**
	(0.04)	(0.05)	(1.82)	(1.55)
Average University: $Grade_{i,j,k,t_{-3}}$	0.60**			
	(0.03)			
t_0 Group assignment date Fixed Effects	Yes	Yes	Yes	Yes
$t_0\times$ Major j Fixed Effects	Yes	Yes	Yes	Yes
$t_0\times$ Major j $\times$ Study progress Fixed Effects	No	Yes	No	No
Adj. R^2	0.71	0.28	0.29	0.23
N	521	763	353	372
# Peer groups	187	209	128	166

Robust *SE* in parentheses; * $p < 0.05$, ** $p < 0.01$

Notes: In columns (1) and (2), the dependent variable is the average university grade of student i in major j at time t_1 (which is after group assignment). In column (2), we add study progress fixed effects. They ensure that the average university grade of students, who complete the group assignment at later stages of their studies, are comparable to those who complete it early on. In column (3), the dependent variable is the Microeconomics I exam result in points of student i in major j at time t_1-t_7. The students considered in column (3) may have written the exam at different dates but always after group assignment. In column (4), the dependent variable is the Microeconomics I exam result in points of student i in major j at time t_{-1}-t_{-3}. In this case, the students may have still written the exam at different dates but always before group assignment. Note that for the exam points more points indicate a better result whereas the best grade is 1.0 and the worst 5.0. k determines the group into which student i is randomly allocated for the group assignment. We include (but do not show) a gender dummy and country of origin dummies in all regressions.

A.4. Other peer network structures

We study different peer network structures. We identify the potential *shining light* of the group as the individual with the best result based on the University Entrance Grade. As shown in Table A.4 column (1), in this setting we cannot observe a significant peer effect. We can contrast this analysis with a potential *bad apple* effect from the individual with the worst University Entrance Grade. Here, we find a significant positive effect (column (2)). This result passes the usual randomness check (column (4)) and also holds in the sample employed for the randomness check (column (3)). This result does not hold if we relax the assumption of individuals also being within the same major j (column (5)). The *bad apple* effect is less robust than the overall finding. Yet, it is stronger than the *shining light* effect, showing that the worst individuals exert a stronger impact on the rest of the group than the best within the group.

Table A.4: OLS estimation with alternative peer measures – Star network.

	Best	Worst		Randomn. worst	Best all j	Worst all j
	$Grade_{i,j,k,t_1}$			$Grade_{i,j,k,t_{-3}}$	$Grade_{i,j,k,t_1}$	
	(1)	(2)	(3)	(4)	(5)	(6)
Peer measures (based on Univ. entrance grade)						
Best$_{t_{-1},j}$ peer: $Peer_{j,k,\tilde{t}_{\leq -1}}$	0.05					
	(0.04)					
Worst$_{t_{-1},j}$ peer: $Peer_{j,k,\tilde{t}_{\leq -1}}$		0.10*	0.12*	0.08		
		(0.05)	(0.06)	(0.08)		
Best$_{t_{-1}}$ peer: $Peer_{j,k,\tilde{t}_{\leq -1}}$					0.01	
					(0.04)	
Worst$_{t_{-1}}$ peer: $Peer_{j,k,\tilde{t}_{\leq -1}}$						0.06
						(0.04)
Control for individual i's ability						
University entrance: $Grade_{i,j,k,\tilde{t}_{\leq -1}}$	0.32**	0.51**	0.54**	0.69**	0.35**	0.51**
	(0.04)	(0.04)	(0.05)	(0.06)	(0.04)	(0.04)
t_0 Group assignment date Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
$t_0\times$ Major j Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.21	0.30	0.37	0.34	0.19	0.27
N	511	520	342	342	589	603
# Peer groups	200	200	170	170	207	206

Robust *SE* in parentheses; * $p < 0.05$, ** $p < 0.01$

Notes: In column (1)-(3) and (5)-(6), the dependent variable is the average university grade of student i in major j at time t_1 (which is after group assignment). In column (4), the dependent variable is the same but at time t_{-3} (before group assignment). The peer measures are based on the university entrance grade at $\tilde{t}_{\leq -1}$ (which was before going to University [i.e., even before t_{-1} or t_{-3}]). k determines the group into which student i is randomly allocated for the group assignment. We include (but do not show) a gender dummy and country of origin dummies in all regressions.

B. Details for orientation week

We contrast the mandatory group work with an informal social *orientation week*. The properties of the orientation week sample are reported in Table B.5 and B.6.

Table B.5: Shares by (engineering) major and gender.

Engineering Major	Male	Female	Total
Electrical Engineering (in %)	16.6	16.8	16.6
Civil Engineering (in %)	10.7	25.2	13.8
Computer Science (in %)	19.0	14.9	18.1
Mechanical Engineering (in %)	53.7	43.1	51.4
Total (in %)	100.0	100.0	100.0
	(730)	(202)	(932)

Notes: All students take the same business, economics and law classes. They only have different engineering majors.

Table B.6: Descriptive statistics of baseline data set for peer group formation during the orientation week (the very first week of studies).

	Mean	Std. Dev.	Min.	Max.
Average University: $Grade_{i,j,k,\tilde{t}_1}$	3	.81	1	5
University entrance: $Grade_{i,j,k,\tilde{t}_{\leq -1}}$	1.9	.46	1	3.7
Female	.22	.41	0	1
Peer measure				
Next best$_{\tilde{t}_{\leq -1},j}$ peer: $Peer_{i,j,k,\tilde{t}_{\leq -1}}$	1.9	.4	1	3.5
Mean peers: $Peer^{mean_j}_{k,\tilde{t}_{\leq -1}}$	1.9	.23	1.1	2.7
N	932			

Notes: 932 observations come from 932 students i, who first took their general university entrance examination (at a high school), then, in their first week of studies, they took part in the group assignment with random peers. The students took part in the group assignment at 4 different dates between 2011 and 2015. They split up into 4 engineering majors j (see Table B.5). 94.4% of the sample are Germans. The rest have one out of 21 other nationalities.