

EUR Research Information Portal

What drives ability peer effects?

Published in:

European Economic Review

Publication status and date:

Published: 01/07/2021

DOI (link to publisher):

[10.1016/j.euroecorev.2021.103763](https://doi.org/10.1016/j.euroecorev.2021.103763)

Document Version

Publisher's PDF, also known as Version of record

Document License/Available under:

Article 25fa Dutch Copyright Act

Citation for the published version (APA):

Coveney, M., & Oosterveen, M. (2021). What drives ability peer effects? *European Economic Review*, 136, Article 103763. <https://doi.org/10.1016/j.euroecorev.2021.103763>

[Link to publication on the EUR Research Information Portal](#)

Terms and Conditions of Use

Except as permitted by the applicable copyright law, you may not reproduce or make this material available to any third party without the prior written permission from the copyright holder(s). Copyright law allows the following uses of this material without prior permission:

- you may download, save and print a copy of this material for your personal use only;
- you may share the EUR portal link to this material.

In case the material is published with an open access license (e.g. a Creative Commons (CC) license), other uses may be allowed. Please check the terms and conditions of the specific license.

Take-down policy

If you believe that this material infringes your copyright and/or any other intellectual property rights, you may request its removal by contacting us at the following email address: openaccess.library@eur.nl. Please provide us with all the relevant information, including the reasons why you believe any of your rights have been infringed. In case of a legitimate complaint, we will make the material inaccessible and/or remove it from the website.



What drives ability peer effects?[☆]

Max Coveney^{a,*}, Matthijs Oosterveen^b

^a Department of Applied Economics, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam 3062PA, the Netherlands

^b Advance/CSG, Lisbon School of Economics and Management, University of Lisbon, Lisbon 1200-781, Portugal



ARTICLE INFO

Article history:

Received 12 September 2020

Revised 4 March 2021

Accepted 30 April 2021

Available online 11 May 2021

JEL classification:

I23

I28

Z13

D62

Keywords:

Peer effects

Channels

Social interaction

Higher education

ABSTRACT

This paper analyzes the importance of one commonly proposed mechanism potentially driving ability peer effects in the classroom: peer-to-peer social interaction. At a large university, first-year students are randomly assigned to a year-long tutorial group and to one of two subgroups within their tutorial group. The university encourages social interaction within, and not between, these subgroups at the start of the year. Hence, each student can divide her tutorial peers into *close* and *distant* peers. We find spillovers on student performance originating from close peers only. Distant peers are unimportant. This implies that peer-to-peer social interactions drive peer effects.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Economists' ongoing interest in classroom peer effects is not hard to justify; simply by reorganizing peer groups, and without additional resources, it may be possible to increase aggregate student performance. Taking into account important methodological advances (Manski, 1993), the past decade of empirical research includes many well-identified studies in primary, secondary, and tertiary education (Sacerdote, 2014). While these studies have to a large extent confirmed the existence of peer effects in the classroom, little to no evidence exists on the mechanisms through which these effects operate.

This lack of evidence on peer effect mechanisms greatly limits the scope for implementing policies that effectively exploit such effects, the design of which will inevitably depend on the mechanisms at play. As a recent literature review concluded, "we do not yet know enough about the nature of peer effects to engage in social engineering of peer groups to affect students' outcomes in a desired direction" (Sacerdote, 2014, pg. 269). By exploiting random group assignment, this paper analyzes the importance of one commonly proposed and policy-relevant mechanism potentially driving classroom peer effects: peer-to-peer social interaction.

[☆] The dataset used in this research is archived on the [Erasmus University Rotterdam figshare portal](https://figshare.com). The authors have no relevant or material financial interests that relate to the research described in this paper. We thank Adam Booij, John Cawley, Eddy van Doorslaer, Pilar García-Gómez, Hans van Kippersluis, Edwin Leuven, Hessel Oosterbeek, Tom van Ourti, and Dinand Webbink for helpful comments and suggestions. The paper has also benefited from the comments and suggestions of participants at various seminars. Oosterveen acknowledges financial support from the FCT – Fundação para a Ciência e Tecnologia (grant PTDC/EGE-OGE/28603/2017). All omissions and errors are our own.

* Corresponding author.

E-mail addresses: coveney@ese.eur.nl (M. Coveney), oosterveen@iseg.ulisboa.pt (M. Oosterveen).

Ability peer effects encompass any externality through which a classroom peer's ability affects a student's outcome (Sacerdote, 2011). All plausible mechanisms for these externalities can be partitioned into two groups, each with very different stories for why peer effects occur; those that do and those that do not rely on peer-to-peer social interaction.

The basic logic behind explanations relying on social interaction is that peers influence student outcomes via peer-to-peer teaching or discussions of course material, either inside or outside of class. Studies which attribute their results to such mechanisms include Carrell et al. (2009), Booij et al. (2017) and Feld and Zölitz (2017).¹

The other class of proposed explanations do not rely on direct social interaction between students for spillovers to occur. Rather, peers are said to influence outcomes by changing teacher's pedagogical practices (Duflo et al., 2011), causing or reducing classroom disruptions (Lavy et al., 2012; Lavy and Schlosser, 2011), acting as a classroom role model (Hoxby and Weingarth, 2005), or otherwise changing the broader classroom environment.² All of these proposed mechanisms would lead to classroom ability spillovers independent of the level of social interaction between students.

We take advantage of a unique institutional quirk in our educational setting that, in essence, encouraged students to pursue social interaction with one random subset of classroom peers and not with the remaining other subset. We are able to analyze the importance of peer-to-peer social interaction for peer effects by comparing spillovers originating from the two subsets of classroom peers. Intuitively, as the two sets of classroom peers differ only by the level of social interaction a student experiences with them, differences in spillovers between the two groups reflect the importance of the social interaction mechanism.

Our context is the first year of an economics undergraduate program at a large public university in the Netherlands across six cohorts. With the purpose of providing incoming first-year students an opportunity to generate social ties, bonds, and friendships, this university manipulates the social interaction between students and their classroom peers. Students are randomly assigned to a year-long tutorial group of approximately 26 students and to one of two subgroups of 13 students within their tutorial group. The university stimulates social interaction within, and not between, these subgroups during the first weeks of the academic year via several informal meetings. From the perspective of one student, their *close* peers are the subset of their tutorial peers with whom social interaction is encouraged, whereas their *distant* peers belong to the adjacent subset with whom social interaction is not encouraged. For each student, her close and distant peers together form her tutorial group whom she follows classes with throughout the first year. We will use high school GPA as a proxy for own and peer ability, which is shown to be a comprehensive measure of peer quality (Stinebrickner and Stinebrickner, 2006), and allows us to avoid problems related to reflection and common shocks.

We provide two pieces of empirical evidence that show the institutional manipulation of social interaction was successful. Firstly, using first-year tutorial attendance data at the session level, we show that students attend more tutorial sessions with their close peers than their distant peers. Secondly, in contrast to the first year, second-year students can freely choose which tutorial group to register to. We analyze students' registration decisions and show that, in their second year, students are more likely to co-register with their close peers than with their distant peers. The results also show the presence of homophily, which serves to validate group registration as a measure of social interaction. The evidence thus suggests the labels close and distant are appropriate; a student socially interacted significantly more with her close peers than her distant peers.

Exploiting the within-classroom random assignment to these distinct peer groups, we find that classroom peer effects can solely be attributed to a student's close peers. We find no role for distant peers. The point estimates from our linear model imply that a one standard deviation increase in close (distant) peer GPA causes student performance to increase with 0.024 (0.00) standard deviations. Intuitively, as the degree of social interaction is the only difference between a student's close and distant peers, this result implies that classroom peer effects depend on the existence of meaningful social interaction between peers.

Though our baseline results demonstrate that peer effects depend on peer-to-peer social interaction, they do not reveal the precise manner in which such interaction creates spillovers. To investigate this, we use student evaluations and provide suggestive evidence that students with better close peers change their study behavior by substituting lecture attendance for self-study, potentially collaborative with their close peers. We also examine heterogeneity in spillovers by ability, and find that spillovers are roughly linear; high and low ability students seem to benefit (suffer) from social interaction with high (low) ability close peers.

A further contribution of this paper is the introduction of a stylized decomposition of the standard peer effect estimate. This serves two purposes. Firstly, it formalizes the interpretation of our results, and clarifies the manner in which we are able to identify the importance of social interaction. Secondly, by showing how peer effect estimates can be understood as a function which includes the degree of social interaction experienced between peers, it may shed new light on an unresolved puzzle in the literature; the large heterogeneity in the magnitude of peer effect estimates between studies (Sacerdote, 2014). This heterogeneity may simply reflect the degree to which social interaction was present, or perhaps even encouraged, in each context.

¹ Also included in this category are mechanisms like grade competition, or adaptation of academically beneficial habits, within social groups. Other papers that attribute their results to the social interaction channel include Garlick (2018); Brunello et al. (2010); Carrell et al. (2009); Stinebrickner and Stinebrickner (2006); Arcidiacono and Nicholson (2005).

² Other research relying on similar explanations are Oosterbeek and Van Ewijk (2014); Burke and Sass (2013); Lyle (2009); Foster (2006).

Two strategies have been used thus far in the empirical literature to explore channels. The first is to analyze additional outcomes found in secondary data sources, such as course evaluations and surveys.³ The second is to search for heterogeneity in the spillovers by student characteristics that may lend support to a certain mechanism.⁴ While our approach is most in line with the latter strategy, it holds significant advantages over both. Without relying on auxiliary data, we are able to compare the influence of two types of peers within one classroom differing in only a single, mechanism-relevant, and administratively-assigned characteristic; the degree of social interaction.⁵

In addition to classroom peers, the literature has studied different peer groups in various contexts, such as entire schools (Lavy and Schlosser, 2011), dorms (Garlick, 2018) and dorm roommates (Sacerdote, 2001; Zimmerman, 2003), military squadrons (Jones and Kofoed, 2020; Carrell et al., 2013), students in the same group during university orientation week (Thiemann, 2017), students that share more than a certain number of classes (De Giorgi et al., 2010), students who sit next to each other in class (Lu and Anderson, 2014; Hong and Lee, 2017), and university friends (Foster, 2006). It may be that peer effects in different settings and groups work through different mechanisms (Brady et al., 2017). Our focus, however, is on the mechanisms driving ability peer effects in the classroom. Classroom peer effects form the largest part of the ability spillovers literature, and have the greatest potential application in policy given the frequency with which administrators at all educational levels and institutions have discretion over classroom assignment.

Our results have implications for policy makers aiming to exploit peer effects. As it stands, attempts to implement alternative group assignment policies using estimates of peer effects under one particular assignment policy were unsuccessful. A well-known example of this is the study by Carrell et al. (2013), in which the authors use credible estimates of spillovers obtained from ability mixing to construct “optimal” peer groups at the United States Air Force Academy. They find that low ability students whom they intended to help with this group assignment policy actually performed worse than untreated low ability students. The importance of peer-to-peer social interactions for classroom peer effects may provide an explanation for the difficulty of exploiting peer effects; if the patterns and degrees of social interaction differ across group configurations, then peer effects under ability mixing may not provide an accurate description of what will occur under alternative assignment policies.

2. Context

2.1. Institutional setting

Our setting for studying peer effects is the economics undergraduate program at a large public university in the Netherlands. Every year the economics program admits approximately 400 newly enrolled first-year students. During the first two undergraduate years the program is identical for every student, as they follow the same twenty courses across the two years, covering basic economics, business economics, and econometrics. Come the third year, students must choose their own courses. The program only admits Dutch students. The admission requirement is based on a having a pre-scientific high school diploma.

The three academic years are divided into five blocks of eight weeks each (seven weeks of teaching and one week of exams).⁶ Students in the first and second year have one light and one heavy course per block, for which they can earn four and eight credits respectively. Sixty credits account for a full year of study.⁷ Grading is done on a scale that ranges from 1 to 10. Students fail a course if their grade is below 5.5.

In the first and second year, courses consist of both lectures and tutorial sessions. The heavy courses have two to three large-scale lectures per week, while light courses have one to two. Heavy courses have two small-scale tutorials per week, while light courses have one. Lectures and tutorials both last for 1 hour and 45 minutes. While attendance at lectures is voluntary, first-year students have to attend at least 70 percent of the tutorials per course. Students who fail to meet the attendance requirement are not allowed to take the final exam for their course and must wait a full academic year before they can take the course again.

During tutorial sessions a teaching assistant (TA) typically works through question sets based on the materials covered in the lectures. Roughly 10 percent of the TAs are PhDs, with some exceptions the remaining 90 percent are senior students. First-year students are randomly assigned to a tutorial group (see Section 2.3 for details), with whom they follow all the tutorial sessions throughout the whole first year. To verify whether the 70 percent tutorial attendance requirement in first year is met, TAs register attendance at the start of each session. The requirement ensures that students experience a sizable

³ See e.g. Booij et al. (2017); Feld and Zölitz (2017); Lavy et al. (2012); Lavy and Schlosser (2011); Stinebrickner and Stinebrickner (2006)

⁴ See e.g. Brady et al. (2017); Garlick (2018); Oosterbeek and Van Ewijk (2014); Duflo et al. (2011); Brunello et al. (2010); Carrell et al. (2009); Lyle (2009); Foster (2006); Arcidiacono and Nicholson (2005); Hoxby and Weingarth (2005); Hoxby (2000).

⁵ Heterogeneity by typically-used characteristics (such as race in Garlick (2018)) offers only suggestive evidence on mechanisms, and is unable to definitively rule out competing explanations. For instance, own-race peers may be more important both because of higher levels of social interaction, or because own-race students are more likely to be role models (absent any social interaction).

⁶ At the end of the academic year, at the start of summer, there is a resit period. During two weeks first- and second-year students have the opportunity to resit a maximum of three courses.

⁷ In this institution credits are measured through ECTS, which is an abbreviation for European Transfer Credit System. This measure for student performance is used throughout Europe to accommodate the transfer of students and grades between universities. The guidelines are that one ECTS is equivalent to 28 hours of studying.

degree of exposure to tutorials and their tutorial peers, and are not able to voluntarily attend different groups during the first year. We investigate peer effects originating from these first-year tutorial peer groups.

Appendix A.1 gives an overview of the first-year courses, their characteristics, and an accompanying tutorial description. This table shows that most of the courses in first year are (partly) multiple choice and therefore graded without interference by the instructor or TAs. For exams with open questions, instructors typically disallow TAs from grading their own groups. Moreover, this table clarifies that the final grade for most courses is solely determined by individual exams, and not by group assignments within (tutorial) peer groups.

2.2. Close and distant peers

A key institutional feature of the economics program is that each first-year tutorial group is randomly divided into two subgroups. The university stimulates social interaction only within, and not between, these subgroups of students. For each student we term *close* peers to be the subset of their tutorial peers with whom social interaction is stimulated, where *distant* peers are the adjacent subset of peers in the tutorial group with whom interaction is not stimulated. This means that if student i and j are in the same tutorial group but in different subgroups, the close peers of student i will be the distant peers for student j and *vice versa*.

Why does the university create the (close peer) subgroups? The official purpose of these groups is to help first-year students with their transition from high school to university. According to the university, one of the biggest difficulties that students encounter during this transition is the absence of a supportive social network; few, if any, of students' former acquaintances from high school will join them at university.

The university therefore engineers initial interactions between first-year students, providing them with an opportunity to generate such a social network. It does so via compulsory close peer group meetings, primarily via five of them concentrated in the first block of the first year. These meetings revolve around discussion and active student participation, which is why the university creates the smaller subgroups (rather than using the tutorial groups). The subjects and the setting of these meetings are less formal than the tutorial groups. The first close peer group meeting is a get-to-know-you session, where students introduce themselves to the group. The university complements the initial interaction within the subsequent four sessions with a more direct introduction to the study program and university life. As such, the subsequent four sessions consist of group discussions on the use of study timetables, exam preparation, fraud and plagiarism, teamwork, plans concerning the future of their studies, and social activities on campus and in the city. Course material is not discussed during the close peer group meetings. The university assigns senior students as discussion leaders to guide the close peer group meetings. Each meeting lasts for 45 minutes.

The first close peer group meeting is in the first week of university. As well as meeting each other in the subsequent tutorial sessions, which also include the set of distant peers, there are weekly close peer group meetings up until week five. During the first five weeks close peers see each other 20 times; 5 times at the close peer meetings and 15 times at the regular tutorials. There are four remaining meetings with the close peer groups that are evenly spread out across the year (one per block). During these remaining meetings, students evaluate the two courses within that block and discuss their study progress with each other.⁸ An overview of the first block and the whole undergraduate program can be found in Fig. 1.

Given their content and concentration at the start of the first year, the close peer group meetings should be viewed as a vehicle for establishing what is likely the first group of fellow students that a new student will interact with and, more generally, generate social ties, bonds, and friendships with. Section 7 provides empirical evidence for this claim. In particular we show that, while the close peer meetings did not stigmatize distant peers, a student socially interacts significantly more with her close peers than her distant peers.

2.3. Assignment of students to groups

During the final year of students' pre-scientific education, and before the start of the academic year, students must preregister for the economics program. Those who have done so are requested to come to campus on the first day of the academic year to confirm their registration. This is done by means of approximately 10 to 15 administrative personnel, who add students' numbers and names to an electronic register.

A list containing the information of all new first-year students who confirmed their registration is sent to an administrative worker. This list is sorted by a randomly assigned ID and group membership is determined on a rotating basis. The first student on the list is allocated to tutorial group 1, close peer group 1A; the second student is allocated to tutorial group 2, close peer group 2A; the third student is allocated to tutorial group 3, close peer group 3A, and so forth. The allocation continues until the maximum tutorial group has been reached, after which the rotation begins again by allocating the next unassigned student to tutorial group 1, close peer group 1B, the next student to tutorial group 2, close peer group 2B, and

⁸ While the students do not get any credits for the close peer group meetings, according to the Teaching and Examination Regulations students must attend all of these meetings in order to pass the first year. Our administrative attendance data reveals students attend on average 94 percent of the sessions.

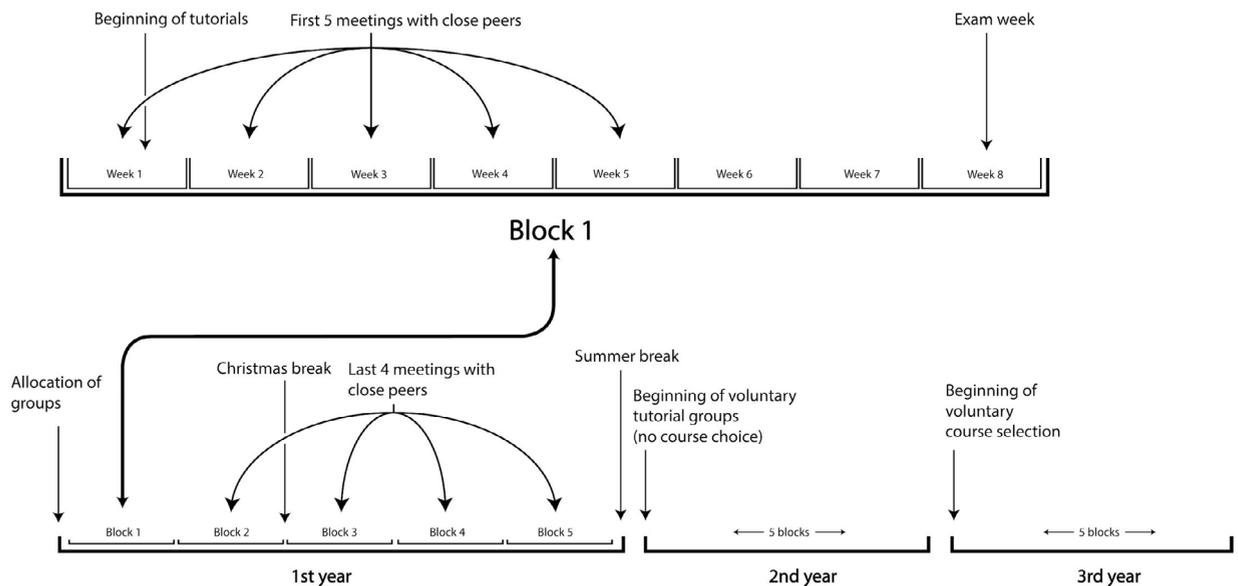


Fig. 1. Overview Undergraduate Economics Program. Notes: This figure gives an overview of the characteristics of the undergraduate Economics program relevant to our study.

so forth. The university uses this allocation method to ensure that students are exposed to new peers and that the groups are roughly of equal size.⁹

Fig. 2 clarifies the structure of tutorial, close, and distant peer groups for a hypothetical cohort. The 144 students, represented by dots, are distributed across 6 tutorial groups and 12 subgroups, two for each tutorial group. The close peer group is defined as the subgroup that the student is assigned to, whereas the distant peer group is defined as the adjacent subgroup in the same tutorial group. For instance, for a student who is assigned to subgroup 1A (1B), the close peer group is subgroup 1A (1B) and the distant peer group is subgroup 1B (1A). Students assigned to subgroups 1A and 1B attend tutorials together as tutorial group 1.

A student who wants to follow the program, but did not show up at the first day of the year, is allocated to a group at the discretion of the administrative worker. Reallocating a student to a different group only happens in case of special circumstances, such as when a student has special needs, is a student athlete, or has some otherwise unresolvable scheduling conflicts. Again, the groups to which these students are reallocated to is at the discretion of the administrator. Our data does not allow us to observe which student registered late or ended up in their group via a reallocation. According to the administrative worker these cases are rare, but may result in slightly different variation in peer ability and class size than would have been observed when strictly following the allocation procedure described above. We present balancing tests in Section 4 that cannot reject the final allocation resulted from a random assignment of students to groups.

3. Data

Our main source of data is the administrative database of the university from the academic years 2009-10 until 2014-15. This database includes all courses followed, and each corresponding grade obtained, by every student for these six years. Additionally this database includes a rich set of student characteristics; student identification number, gender, age, residential address, immigrant status, high school GPA and zip code, and the groups students have been assigned to in their first year.¹⁰ Student identification numbers are increasing in tenure and determined by the time of preregistration. We extended this administrative database with additional records of student outcomes and choices at university, described below.

⁹ We conducted several interviews with the administrative worker and university administrators, and received accompanying documentation, in order to confirm that the allocation process occurred as described. The same administrative worker has been in charge of this process across the six cohorts we study. The allocation process is done with BusinessObjects BI and Microsoft Excel software.

¹⁰ Following the definitions of the Dutch Statistics Office, we classify a student as having a non-western immigrant background if at least one of their parents was born in a so-called non-western country, primarily Suriname, Morocco, and Turkey. All other students with an alternative immigration origin are classified as having a western immigrant background. Students born in the Netherlands to Dutch parents are classified as having a native Dutch background.

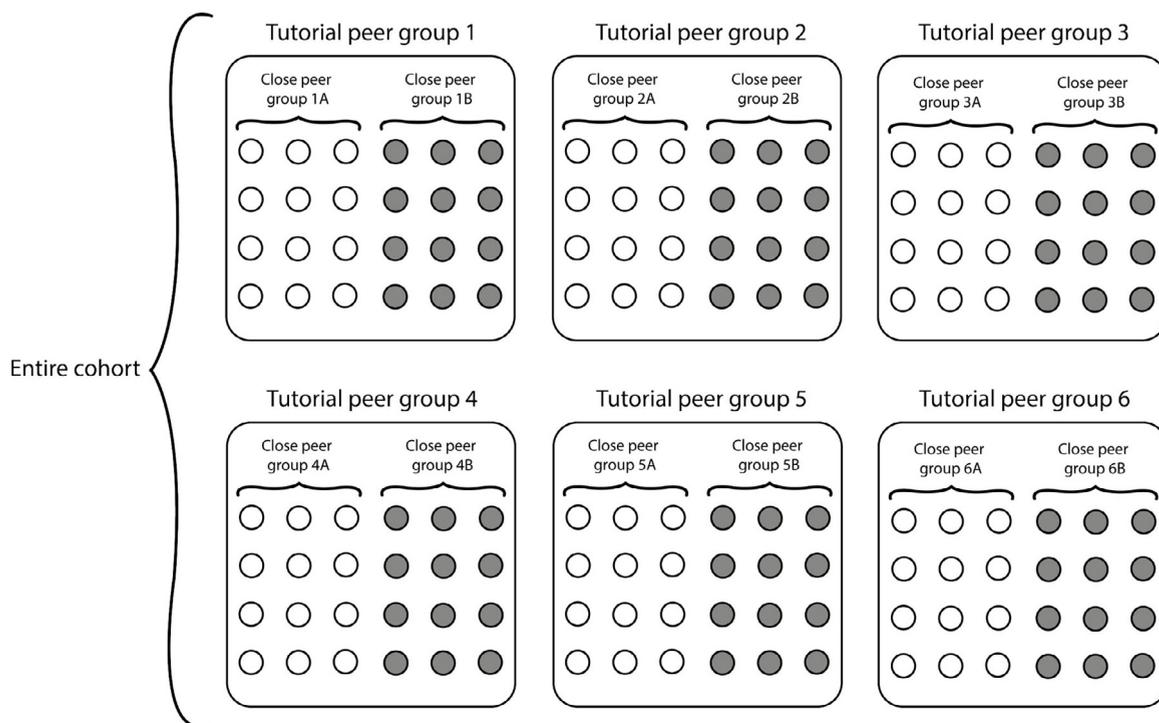


Fig. 2. Structure of Tutorial and Close Peer Groups. Notes: This figure gives a graphical representation of the allocation to tutorial and close peer groups for a hypothetical cohort.

Our analysis is based on almost 19,000 first-year grades from 2,300 students across six cohorts.¹¹ This sample only includes a student's first attempt at completing a course. Although we also observe resits, which are taken at the end of the academic year at start of summer, we exclude them from our analysis as they do not require preparation via tutorials.

High school GPA is a 50-50 weighted average of grades obtained during the last three years of high school and on the nationwide standardized exams at the end of high school across all courses. We use high school GPA as a proxy for the ability of students and their peers. In case of classical measurement error, our estimate for spillovers would be attenuated as students are randomized into groups (Feld and Zölitz, 2017).¹²

3.1. Attendance and student evaluations

In first year all students are required to attend at least 70 percent of the tutorials per course. Attendance is registered by TAs during each session, uploaded to the university portal, and used to verify that the attendance requirement is met. We merge this attendance data with the administrative database, which allows us to observe attendance at the student-course-session level for 98.5 percent of the student-course observations.¹³

At the end of the course, students are invited by email to fill in student evaluations. A set of 20 questions are asked covering 9 characteristics of the course, which are detailed in Appendix A.2. Merging the student evaluations to the administrative data gives a response rate of roughly 30 percent. Column (1) of Appendix A.8 reveals that participating in the course evaluations is selective; students with a better high school GPA are more likely to respond. However, column (1) also shows the absence of a relationship between the high school GPA of students' close and distant peers and their response

¹¹ This sample excludes some students. For 227 students we do not observe high school GPA (225 students) or one of the main control variables (2 students). Furthermore, to ensure that peer GPA consists of an appropriate number of students, we dropped fourteen tutorial groups (215 students) for whom we observe less than ten students' GPA in at least one of the two (close peer) subgroups. Note that these groups occurred because of missing data on high school GPA and because some students were reallocated after the initial assignment. Our results are robust to the inclusion of these groups.

¹² There are two potential sources of measurement error in our measure of ability. First, we do not separately observe the grades from the unstandardized school exams and from the nationwide standardized exams. It should be noted, however, that the Dutch Inspectorate of Education pays strong attention to schools where the grades on school exams deviate more than 0.5 points from grades on the nationwide standardized exams (DUO, 2014). Second, although students have followed the same level of education in high school (pre-scientific), entering the last three years of high school students must choose one of four tracks. Though these tracks share compulsory courses, some courses between tracks differ. For a subsample we can show that over 70 percent of our students followed the same track.

¹³ For our grade-analysis we use the whole sample. Results are identical for the sample that is matched to the attendance data. We verified that peer high school GPA cannot explain whether a student is matched.

rate. Results using the course evaluations should be interpreted with caution, and we use them to provide supplementary evidence on the channels of peer influence.

3.2. Tutorial and working group registration

In contrast to the first year, in the second year all students have to register for the tutorial groups themselves. Two to three weeks before the start of each block, second-year students receive an email asking them to register for a tutorial group for each of the two courses. The registration takes place via the university portal and students may register freely to any tutorial group that is not already at full capacity. One of the ten second-year courses revolves around writing a research report in groups of three to four students. Instead of registering for a tutorial group, students have to register for these smaller working groups at the start of the course, again via the university portal. For all six cohorts we extend the administrative data with students' second-year tutorial and working group registration.¹⁴ Because acquainted students coordinate their registration decisions to end up in the same groups, we will use these data in Section 7 to analyze whether students co-registered more with their close peers than their distant peers.

3.3. Descriptive statistics

Table 1 shows the descriptive statistics by cohort. Panel A provides an overview of the student characteristics. Panel B does the same for student outcomes. All student characteristics show similar values across cohorts. The percentage of women is approximately 20 percent, the students are on average 19.5 years old halfway into their first year, roughly 25 (5) percent of the students are Non-Western (Western) immigrants, and high-school GPA is close to the nationwide average of 6.7 (scale from 1 to 10). Appendix Figure A.1 shows histograms of student's own high-school GPA, the leave-out mean for the tutorial- and close peer group, and the mean for the distant peer group. Notice that, in contrast to the leave-out mean for the close peer group, the mean for the distant peer group takes upon identical values for everybody in the same subgroup. This explains the somewhat more discrete nature of this figure. A histogram of the leave-in mean for the close peer group is similar to the mean for the distant peer group.¹⁵

Table 1 further shows that the size of the close peer group fluctuates between 11 and 18 students across cohorts. In 2009 the groups were somewhat larger due to an unexpectedly high number of enrolled students. University grades seem to gradually increase, also reflected by the increase in the number of credits earned. This is most likely the consequence of stricter academic dismissal policies introduced halfway in our sample. Course dropout occurs if a student does not attend the final exam for that particular course. Across cohorts, 8 to 19 percent of the students dropped out of both courses in block 5, the final block of the first year. We refer to this as student dropout.

4. Empirical specification

To derive our empirical model we start with the linear-in-means specification commonly used in the peer effects literature:

$$Y_{igct} = \beta_0 + \beta_1 \overline{GPA}_{(-i)g} + \beta_2 GPA_i + \epsilon_{igct}, \quad (1)$$

where Y_{igct} is the grade at university of student i from cohort t on course c , who is assigned to tutorial group g . GPA_i is i 's average grade obtained in high school, $\overline{GPA}_{(-i)g}$ is the leave-out mean of high school GPA of tutorial group g for student i , and ϵ_{igct} is a random variable reflecting unobserved determinants of a student's grade.

The institutional manipulation of the degree of social interaction within, and not between, subgroups of the larger tutorial groups allows us to extend this standard model. We can make a distinction between the leave-out mean of the close peer group $\overline{GPA}_{Close(-i)g}$ (the subgroup to which a student is assigned) and the mean of the distant peer group $\overline{GPA}_{Distant_g}$ (the adjacent subgroup in the same tutorial group), and replace $\overline{GPA}_{(-i)g}$ in Eq. 1 by the following expression:

$$\overline{GPA}_{(-i)g} = \frac{N^C - 1}{N^C + N^D - 1} \overline{GPA}_{Close(-i)g} + \frac{N^D}{N^C + N^D - 1} \overline{GPA}_{Distant_g},$$

where N^C and N^D are the total number of students in the two subgroups within a tutorial group. In practice, $N^C \approx N^D \approx 13$. This substitution allows us to arrive at the following specification:

$$Y_{igct} = \beta_0 + \beta_1^C \overline{GPA}_{Close(-i)g} + \beta_1^D \overline{GPA}_{Distant_g} + \beta_2 GPA_i + \epsilon_{igct}. \quad (2)$$

¹⁴ For the final cohort in our data, i.e., the 2014-15 cohort, we are not able to observe the second-year grades (in the year 2015-16). For this cohort we merge the group-registration data on the student level and we take into account all students that registered for a group (even if they did not obtain a valid grade for the course).

¹⁵ Angrist (2014) shows that using leave-in means, rather than leave-out means, would only change the peer-effects estimate for close peer high school GPA by a factor of $N^C / (N^C - 1)$, where N^C is the size of the close peer group. Therefore, this distinction has little bearing on our results.

Eq. 2 tests the restriction imposed by Eq. 1 that the spillovers β_1 from close and distant peers are identical.¹⁶ How exactly should we interpret differences between β_1^C and β_1^D , and under what conditions does this allow us to identify the role of social interaction for peer effects?

4.1. Stylized decomposition

To illustrate how one should interpret differences between β_1^C and β_1^D , we adopt a stylized decomposition by splitting the composite classroom spillover coefficient into two components: spillovers that depend on meaningful peer-to-peer social interaction (α_1^{SI}), and spillovers that work independent of such interaction (α_1^{non-SI}). In particular, we express the peer effects coefficients from Eq. 2 as:

$$\begin{aligned}\beta_1^C &= s^C \times \alpha_1^{SI} + \alpha_1^{non-SI}, \\ \beta_1^D &= s^D \times \alpha_1^{SI} + \alpha_1^{non-SI},\end{aligned}$$

where the variable s^C (s^D) is some measure of the degree of meaningful social interaction between a student and her whole close peer group (whole distant peer group). A high s thus reflects a strong degree of social interaction with all students in the peer group, or potentially with a large enough random subsample of students. It will be low, however, when a student only interacts with a small self-selected subsample of students from the peer group, or does not interact with students from the peer group at all. To ease interpretation, let s be measured on a scale that equals 1 for the close peer group, *i.e.*, $s^C = 1$. In this case s^D , which we will conveniently refer to as s , reflects the depressed level of meaningful social interaction between distant peers compared to close peers. The difference between β_1^C and β_1^D can then be written as:

$$\beta_1^C - \beta_1^D = \alpha_1^{SI} \times (1 - s).$$

In Section 7 we show that the close peer group meetings caused students to experience a higher (lower) degree of social interaction with their close (distant) peers, *i.e.*, we show that $1 > s \geq 0$. Applying these bounds to the expression above, a finding of $\beta_1^C > (\beta_1^C - \beta_1^D) > 0$ would allow us to conclude that peer-to-peer social interactions are an important driver of classroom peer effects, *i.e.*, $\alpha_1^{SI} > 0$. Note that, as the value for s is generally unknown, the exact magnitude of this importance cannot be established in most cases. However, as we will show when interpreting our baseline results, our particular set of findings are informative about the value of s , and thus allow more precise statements about the magnitude of α_1^{SI} and α_1^{non-SI} .

The stylized decomposition clarifies the two main assumptions that are required to identify the importance of the peer-to-peer social interaction channel. Assumption (1) is that $\hat{\beta}_1^C$ and $\hat{\beta}_1^D$ are unbiased estimates for β_1^C and β_1^D . This implies that the covariance between close and distant peer high school GPA and ϵ_{igct} must be zero. The validity of this assumption will be tested in the proceeding subsection.

Assumption (2) is that β_1^C and β_1^D can be described by identical functions that only differ in the level of s . This assumption would be violated if there exists a direct effect, unrelated to s , on later grades from attending these meetings with high ability peers. Similarly problematic would be a scenario where the close peer group meetings somehow altered the parameters α_1^{SI} and α_1^{non-SI} by subgroup, next to a change in s . The nature of the close peer group meetings makes both scenarios implausible. The meetings are short-lived and informal, designed to engineer initial interactions at the start of the year, and not used to discuss any course material. Given this, the only lasting consequence of these meetings is likely the establishment of a close peer group; *i.e.*, a change in s . Section 7 provides empirical support for this claim.

This decomposition may also shed new light on two unresolved puzzles in the literature. Firstly, as noted by Sacerdote (2014), the magnitude of classroom ability peer effect estimates appears to vary widely between studies. This decomposition demonstrates that even if the parameters α_1^{SI} and α_1^{non-SI} are reasonably stable across studies, differences in s can nevertheless produce heterogeneity in peer effect estimates.

Secondly, as it stands, attempts to implement alternative group assignment policies using estimates of peer effects under one particular assignment policy were unsuccessful. A well-known example of this is the study by Carrell et al. (2013). Contrary to their prediction, based on estimates from ability mixing classrooms, Carrell et al. (2013) find that the grades of low-ability students deliberately placed in (bimodal) classrooms with only high-ability students decreased compared to low-ability students in control (ability mixing) classrooms. One explanation for this puzzling result is that s depends upon the group assignment policy that is considered, where in Carrell et al. (2013) s might have been lower in the bimodal classrooms than in the control classrooms.

4.2. Balancing tests

Are the peer effects estimates from Eq. 1 and (2) unbiased? As the models use a predefined measure for peer ability, we avoid both the reflection problem and contamination from common shocks. The main identifying assumption, however, is that peer high school GPA is uncorrelated with other characteristics that might determine a student's grade; the covariance

¹⁶ In fact, because the mean GPA from the distant peer group contains one more student than the leave-out mean of the close peer group, if the spillovers from close and distant peers are identical then $\beta_1^C = \beta_1^D (\frac{12}{13})$.

Table 1
Descriptive Statistics per Cohort.

	2009-10 Mean (SD)	2010-11 Mean (SD)	2011-12 Mean (SD)	2012-13 Mean (SD)	2013-14 Mean (SD)	2014-15 Mean (SD)
Panel A: Student Characteristics						
Female	0.21 (0.41)	0.21 (0.41)	0.22 (0.42)	0.22 (0.42)	0.21 (0.40)	0.23 (0.42)
Age	19.54 (1.65)	19.62 (1.29)	19.61 (1.28)	19.57 (1.57)	19.67 (1.34)	19.48 (1.42)
Distance to University (km)	21.77 (26.37)	24.03 (30.96)	21.62 (26.20)	22.56 (26.32)	26.39 (31.96)	18.08 (20.32)
Non-Western Immigrants (%)	0.25 (0.43)	0.29 (0.45)	0.24 (0.43)	0.21 (0.41)	0.25 (0.43)	0.22 (0.41)
Western Immigrants (%)	0.08 (0.27)	0.05 (0.23)	0.06 (0.23)	0.08 (0.27)	0.05 (0.22)	0.06 (0.23)
Own High School GPA	6.72 (0.54)	6.60 (0.48)	6.63 (0.49)	6.62 (0.47)	6.68 (0.56)	6.68 (0.47)
Tutorial High School GPA	6.72 (0.09)	6.60 (0.10)	6.63 (0.10)	6.62 (0.10)	6.68 (0.13)	6.68 (0.09)
Close Peer High School GPA	6.72 (0.12)	6.60 (0.12)	6.63 (0.16)	6.62 (0.13)	6.68 (0.17)	6.68 (0.14)
Distant Peer High School GPA	6.73 (0.11)	6.60 (0.11)	6.63 (0.15)	6.62 (0.13)	6.69 (0.17)	6.68 (0.14)
Tutorial Group Size	35.30 (1.52)	26.84 (2.80)	22.31 (1.15)	22.08 (1.32)	26.12 (1.29)	24.17 (1.63)
Close Peer Group Size	17.72 (1.35)	13.51 (1.85)	11.19 (0.88)	11.08 (0.94)	13.12 (1.10)	12.12 (1.06)
Number of Students	458	371	356	308	442	361
Panel B: Student Outcomes						
Grades	5.98 (1.76)	5.91 (1.71)	6.38 (1.55)	6.06 (1.64)	6.21 (1.68)	6.35 (1.41)
Attendance	0.89 (0.16)	0.89 (0.12)	0.89 (0.10)	0.88 (0.11)	0.88 (0.11)	0.89 (0.10)
Number of Student-Grades Obs.	3598	2999	3098	2580	3462	2999
Number of Student-Att. Obs.	3433	2955	3094	2577	3436	2950
Number of Credits per Student	29.74 (19.62)	30.06 (18.88)	37.96 (18.12)	33.53 (20.59)	32.39 (21.61)	40.00 (20.69)
Number of Courses per Student	8.49 (2.48)	8.79 (2.26)	9.29 (1.84)	8.76 (2.25)	8.43 (2.55)	8.94 (2.24)
Dropout	0.18 (0.38)	0.16 (0.37)	0.08 (0.27)	0.15 (0.36)	0.19 (0.39)	0.12 (0.33)

Notes:

1. Table shows the mean and standard deviation per cohort of student characteristics (Panel A) and student outcomes (Panel B). Panel B is further divided into student-course level outcomes (first section) and student level outcomes (second section).
2. Age is evaluated on January 1st in the academic year that the cohort started. Distance to University refers to the number of kilometers from a student's registered address to the university. High school GPA and university grades are unstandardized, measured on a scale from 1 to 10.
3. Dropout is the fraction of students who did not write an exam in the last block of the first year (block 5).

between peer high school GPA and ϵ_{igt} needs to be zero. Random assignment of students to groups makes this identifying assumption likely to hold.

We test this identifying assumption in several ways. First, we analyze whether the treatment, in the form of assigned peer ability, can be explained by background characteristics X_i - student number, gender, age, distance to university, and immigrant status - or high school GPA:

$$\overline{GPA}_{(-i)g} = \gamma_0 + \gamma_1 X_i + \gamma_2 GPA_i + T_t + \epsilon_{igt}.$$

We include cohort fixed effects (T_t) as the assignment into groups takes place cohort-by-cohort. Estimates of γ_1 or γ_2 that are different from zero most likely violate the identifying assumption mentioned above. Table 2 shows the results of this test, where column (1) to (3) take tutorial, close, and distant peer high school GPA as outcome variables respectively. Across the three specifications we find all background characteristics and high school GPA to be individually and jointly insignificant.¹⁷ This stands in stark contrast to the joint significance of student characteristics in a regression where first-year GPA at university is taken as an outcome variable (p -value < 0.000; both with and without the inclusion of high school GPA as a student characteristic).

Our second balancing test is more flexible. We regress background characteristics and high school GPA on close peer group dummies and cohort fixed effects. Next, in a separate model we regress the student characteristics upon cohort fixed effects only and perform a F-test on the small versus big model. This test would reveal if students with certain characteristics cluster together in groups. Appendix Table A.3 shows the F-test does not reject the null hypothesis for all student characteristics. In other words, a small model with cohort fixed effects only is favored above a model that also includes close peer group dummies.

We perform a similar analysis per cohort. We regress each student characteristic on a set of close peer group dummies separately for each cohort. Appendix Figure A.2a plots the histogram of the p -values of the close peer group dummies obtained from these regressions. As expected, the p -values are roughly uniformly distributed, where for instance roughly 10 percent of the p -values are below 0.10. Fig. shows the results for this analysis are identical if close peer group dummies are replaced with tutorial group dummies. A Kolmogorov-Smirnov equality of distribution test does not reject the null-hypothesis of a uniform distribution in both cases; the p -values are equal to 0.73 and 0.78 for the histograms belonging to the close- and tutorial peer group dummies respectively.

¹⁷ If we regress student high school GPA on peer high school GPA we reach identical conclusions. Guryan et al. (2009) argue this balancing test should also control for the mean high school GPA of all peers that can be matched with student i in group g . In our case this control would be the leave-me-out mean GPA of her cohort. This is infeasible as there is no variation in the group that student i can be matched too. Indeed, GPA_i is related to the mean GPA of her cohort \overline{GPA}_t and the leave-me-out mean GPA of her cohort, $\overline{GPA}_{(-i)t}$, by the following identity: $GPA_i = N \times \overline{GPA}_t - (N - 1) \times \overline{GPA}_{(-i)t}$.

Table 2
Balancing Tests for Peer Ability.

	Tutorial Peer GPA (1)	Close Peer GPA (2)	Distant Peer GPA (3)
Student Number	-0.0144 (0.0404)	-0.0192 (0.0449)	-0.0054 (0.0406)
Female	-0.0378 (0.0386)	-0.0289 (0.0441)	-0.0297 (0.0482)
Age	-0.0088 (0.0228)	-0.0006 (0.0237)	-0.0126 (0.0201)
Distance to University	-0.0128 (0.0145)	0.0012 (0.0175)	-0.0214 (0.0153)
Non-Western Immigrant	0.0191 (0.0475)	-0.0251 (0.0450)	0.0503 (0.0501)
Western Immigrant	-0.0152 (0.0695)	-0.0726 (0.0753)	0.0381 (0.0788)
Own GPA	0.0092 (0.0289)	-0.0193 (0.0293)	0.0329 (0.0260)
Observations	2296	2296	2296
Adjusted R ²	0.150	0.084	0.098
F-test	0.32	0.26	0.77
p-value	0.942	0.967	0.617

Notes:

- All regressions also include cohort fixed effects.
- Peer GPA refers to the leave-out mean of high school GPA for the tutorial- and close peer group, and to the mean for the distant peer group. All dependent and independent variables are standardized except for the female dummy, the non-western immigrant dummy, and the western immigrant dummy.
- The F-test, and corresponding p-value, refer to a test for the joint significance of all the independent variables shown in the table.
- Standard errors in parentheses, clustered on the tutorial group level.
- * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3
Peer Effects on First-Year Course Grades.

	Grades (Standardized)				
	(1)	(2)	(3)	(4)	(5)
Tutorial Peer GPA	0.0189 (0.0117)				
Close Peer GPA		0.0243** (0.0105)		0.0242** (0.0107)	0.0243** (0.0110)
Distant Peer GPA			0.0042 (0.0131)	0.0018 (0.0130)	0.0020 (0.0131)
Close × Distant Peer GPA					-0.0143 (0.0117)
Own GPA	0.3351*** (0.0124)	0.3358*** (0.0124)	0.3350*** (0.0123)	0.3358*** (0.0124)	0.3358*** (0.0126)
Observations	18,736	18,736	18,736	18,736	18,736
Adjusted R ²	0.325	0.325	0.324	0.325	0.325

Notes: 1. All regressions include course-cohort fixed effects and controls; student number, a female dummy, age, distance to university, and a non-western and western immigrant dummy. 2. Peer GPA refers to the leave-out mean of high school GPA for the tutorial- and close peer group, and to the mean for the distant peer group. Own GPA refers to own high school GPA. All GPA measures are standardized. 3. Standard errors in parentheses, clustered on the tutorial group level. 4. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Allocation of TAs to tutorial groups is done for each course by the course instructor based upon scheduling restrictions. To confirm this, we made use of our limited TA data. For roughly 90% of the tutorial groups we retrieved the names of the TAs (per course) via the student evaluations. For the remaining 10% of the groups we could not find a TA because none of the students within a group participated in the evaluation. We subsequently (manually) coded their gender and whether he or she was a PhD where possible. If instructors base TA assignment on the difficulty of groups instead, they might, for instance, assign PhD's to low GPA groups. However, regressing TA type on tutorial peer GPA confirms that instructors do not base TA assignment on class composition (see Appendix Table A.4). The same assignment method is used for the discussion leaders that guide the close peer group meetings, though we cannot confirm this empirically as the discussion leaders are not in the student evaluations data.

We conclude that we are able to identify causal peer effects, and estimate Eq. 1 and (2) while also including course-cohort fixed effects and background characteristics as explanatory variables; student number, gender, age, distance to university, and immigrant status. The baseline results are identical when we do not control for background characteristics. We cluster standard errors at the tutorial group level; the residuals are allowed to correlate within a student (between courses) and between students of the same tutorial group. Own GPA, peer GPA, and the outcome variables (when suitable) are standardized over the estimation sample, such that estimates can be interpreted in terms of standard deviations.

5. Baseline results

Before presenting the baseline results, we discuss the extent to which course- and student dropout could potentially bias our estimates. Table 1 showed that the student dropout rate at the end of first year is between 8 and 19 percent across the six cohorts. In Section 5.5 we will show that average peer high school GPA has no impact on the number of courses a student attends the final exam for nor on whether the student dropped out by the end of first year. We can show, but omit for brevity, that these null-results for number of courses and student dropout extend to the non-linear model used in Section 6. Selection bias therefore does not contaminate the following baseline peer effects estimates.

5.1. First-Year grades

Table 3 presents our baseline results, where first-year grades are regressed upon peers' average high school GPA. Column (1) shows the estimated effect of tutorial peers. The positive coefficient has a p -value of 0.11 and shows that a one standard deviation increase in tutorial peer high school GPA increases a students' first-year grade by 0.019 standard deviations. Columns (2) and (3) show the effect while separating the tutorial group by one's close- and distant peers. This reveals that the positive spillovers are entirely driven by close peers. The estimate for peer GPA when moving from tutorial to close peers in column (2) increases somewhat in magnitude and precision. It is statistically significant at the 5%-level. The estimate for distant peers in column (3) is economically and statistically indistinguishable from zero. Column (4) shows that the estimates for close- and distant peer high school GPA are identical to the estimates in columns (2) and (3) when including both in one regression.

In terms of the Dutch grading scale, columns (2) and (4) imply that increasing the close peers' high school GPA from 6.5 to 7 increases a student's grade from 7 to 7.14. This is economically small, but 2.1 times the size of Feld and Zölitz (2017), while Booij et al. (2017) find no peer spillovers in their linear-in-means specification. Both of these studies investigate spillovers in a context similar to ours; a public university in the Netherlands.

Column (5) of Table 3 includes an interaction term between close and distant peer high school GPA. The interaction term might pick up possible complementarities between channels that rely on social interaction *versus* those that do not. For instance, having a superstar (but distant) student in class posing insightful questions may only increase grades if one has high ability (close) peers to discuss the questions with. Alternatively, the interaction term can be interpreted as a more direct test for the presence of teacher response, as teachers might only increase their instruction quality if both close and distant peer GPA are high. However, the results show that the interaction term is insignificant, which supports our discussion of the two classes of channels as separate.

5.2. Interpretation of baseline results

The findings in Table 3 are similar for all remaining analyses: close peers are important for student performance ($\beta_1^C > 0$), while distant peers are unimportant ($\beta_1^D = 0$). Referring to the decomposition in Section 4.1, this finding implies that peer-to-peer social interactions are an important driver of classroom peer effects. Under one additional assumption, however, our particular result allows us to make precise statements about the exact magnitude of this importance.

Following the decomposition, the finding of $\beta_1^C > 0$ and $\beta_1^D = 0$ can be written as:

$$\alpha_1^{SI} + \alpha_1^{non-SI} > 0,$$

$$s \times \alpha_1^{SI} + \alpha_1^{non-SI} = 0.$$

If we additionally assume that α_1^{SI} and α_1^{non-SI} do not have opposite signs, then the restrictions imposed by these expressions imply that $s = \alpha_1^{non-SI} = 0$ and $\beta_1^C = \alpha_1^{SI}$. In words, no meaningful peer-to-peer social interaction occurred between distant peers, no peer effects work independently of meaningful peer-to-peer social interaction, and equivalently all classroom peer effects rely on such meaningful interaction.

This additional assumption rules out the possibility that while spillovers from high ability peers dependent on social interaction increase student performance ($\alpha_1^{SI} > 0$), spillovers from high ability peers that do not depend on social interaction decrease student performance ($\alpha_1^{non-SI} < 0$). How realistic is this assumption? Firstly, from a theoretical point of view, we note that no examples of the two mechanisms provided in the Introduction are consistent with these coefficients having negative signs. Secondly, empirically we find no evidence in the proceeding results - considering robustness checks, alternative outcomes, and non-linear models - of a negative peer effect from being assigned to high ability close or distant peers, which one might expect if α_1^{non-SI} was indeed negative. Given that the independent empirical evidence presented in

Section 7 is also broadly consistent with the notion that $s = 0$, we think this assumption, and the interpretation it implies, most closely matches our empirical results.

5.3. Placebo analyses

The results above use analytic standard errors. In this section we present p -values based on randomization inference. This method involves re-drawing a large number (10,000) of randomly assigned hypothetical tutorial and close peer groups, respecting the size of the original groups. For each of these hypothetical groups, we re-run the models presented in Table 3 in order to assess the effect of the hypothetical peers' high school GPA on students' first-year grades. Comparing the actual estimate to the estimates from the hypothetical groups allows us to test the sharp null hypothesis that peer effects are equal to zero (Athey and Imbens, 2017). The results for the corresponding exact p -values are presented in Appendix Table A.5 and Fig., which are nearly identical to those presented in Table 3.

Additionally, these results may address one of the concerns of Angrist (2014). This paper demonstrates that the peer effects estimate is identical to the (scaled) difference between a 2SLS estimator using peer group dummies as instruments for individual high school GPA and an OLS estimator of individual GPA. In some settings this may lead to a spurious, mechanically driven finding that resembles a peer effect. In our setting, however, with random assignment of students to many small groups, there is little reason for this estimate to be different from zero in the absence of spillovers (Angrist, 2014). This is confirmed by the fact that the peer effect coefficients from the 10,000 hypothetical groups, containing unconnected students, are centered around zero.

We use a similar procedure to make a stricter distinction between close and distant peers. Recall that each tutorial group contains two subgroups, which implies that each student appears both in a close peer group and in a distant peer group. As a robustness check, we run regressions on a sample while only including students from one of the two subgroups within each tutorial group. The measures of distant peer GPA for the included students in the regression remain the same, but the estimation sample does not include their distant peers. Hence, the procedure is to randomly drop one of the two subgroups within each tutorial group, estimate the specifications in columns (2), (3) and (4) of Table 3 (on roughly half the number of observations), and then repeat this 10,000 times. Appendix Figure A.5 presents histograms of the close and distant peer effect estimates for these 10,000 draws. The estimates are heavily centered around our original estimates, as indicated by the dashed red bars, further corroborating the robustness of our results.

5.4. Robustness

We will show below that peer GPA does not affect (course) dropout, which implies that our results are not contaminated by selection bias. However, low GPA students do take fewer courses and have a higher probability of dropping out by the end of first year. This means that a student randomized into a tutorial group with many low ability rather than high ability students will experience a larger amount of course dropout among her peers, and thus have a smaller actual class size. This results in a positive correlation between peer GPA and class size, which could partly explain our baseline results if class size also impacts grades. Appendix Figure A.5 plots the number of students writing the final exam per block as a fraction of the initial students, separately for high, average, and low GPA close peer groups. This reveals that course dropout increases during the year, being 15 to 20 percent at the end of the first year. It also reveals that dropout is somewhat larger for low GPA close peer groups.

We investigate whether our results are robust to class size and course dropout in Table 4, which presents the results of our baseline model while including control variables that measure class size and course dropout. Column (1) includes a dummy for the assigned number of students to the close peer group at the start of the first year, column (3) for the actual number of students that wrote the exam for the course, and column (6) for the difference between the two. The latter is a measure for dropout per course. All three columns reveal a stable estimate for close peer GPA, suggesting that class size and course dropout are unlikely to explain our baseline results.

Columns (2), (4), and (7) of Table 4 include the assigned class size, actual class size, and course dropout as continuous variables, while also including their interaction with close peer GPA. The measures for assigned and actual class size in column (2) and (4) are standardized, while the difference between the two in column (7) is unstandardized. As such, the estimate for close peer GPA in (7) measures the peer effect for groups where there has been no course dropout. Again we find stable estimates for close peer GPA across all three columns. Moreover, we find all interaction terms to be unimportant.

Whereas assigned class size is exogenous, one may have remaining concerns that actual class size is an outcome of close peer GPA. Therefore we report an additional specification in column (5) of Table 4, where we use assigned class size as an instrument for actual class size.¹⁸ Using only the variation in actual class size that originates from the assignment, we find our results to be virtually unchanged.

¹⁸ The variation in assigned class size comes partly from the original allocation and partly from the cases in which the administrator reallocates students across tutorial and close peer groups (see Section 2).

Table 4
Robustness of Baseline Peer Effects.

	Grades (Standardized)						
	OLS (1)	OLS (2)	OLS (3)	OLS (4)	IV (5)	OLS (6)	OLS (7)
Close Peer GPA	0.0242** (0.0113)	0.0239** (0.0108)	0.0259** (0.0109)	0.0275** (0.0111)	0.0274** (0.0113)	0.0246** (0.0107)	0.0364** (0.0151)
Close Peer GPA × Assigned Class Size		-0.0045 (0.0089)					
Close Peer GPA × Actual Class Size				0.0050 (0.0075)	-0.0026 (0.0107)		
Close Peer GPA × (Assigned-Actual)							-0.0051 (0.0054)
Distant Peer GPA	-0.0003 (0.0128)	0.0036 (0.0130)	0.0026 (0.0130)	0.0023 (0.0131)	0.0029 (0.0132)	0.0030 (0.0136)	0.0019 (0.0131)
Own GPA	0.3359*** (0.0125)	0.3360*** (0.0124)	0.3359*** (0.0124)	0.3362*** (0.0124)	0.3362*** (0.0123)	0.3357*** (0.0124)	0.3358*** (0.0124)
Observations	18,736	18,736	18,736	18,736	18,736	18,736	18,736
Adjusted R ²	0.326	0.325	0.325	0.325	0.325	0.325	0.325
F-tests on					Assigned Class Size: 130.87		
Excl. Instruments					Its Interaction with Peer GPA: 483.12		
Class-Size Dummies	Yes	No	Yes	No	No	Yes	No
Robustness Check	Assigned Class Size		Actual Class Size		(Assigned-Actual)		

Notes: 1. All regressions include course-cohort fixed effects and controls; student number, a female dummy, age, distance to university, and a non-western and western immigrant dummy. 2. Peer GPA refers to the leave-out mean of high school GPA for the close peer group and to the mean for the distant peer group. Own GPA refers to own high school GPA. All GPA measures are standardized. 3. Column (1) includes dummies for the number of students at the beginning of the year in the close peer group (assigned class size), column (3) for the number of students that wrote the exam for the course (actual class size on the course-cohort level), and column (6) for the difference between these two (assigned-actual). The latter is a measure for course dropout. 4. Assigned and actual class size are standardized in column (2), (4) and (5). The difference between the two in column (7) is not standardized. The coefficient on close peer GPA in column (7) measures spillovers in classes where there has been no course dropout (assigned-actual=0). Roughly 20 percent of the groups experience no course dropout and have a value of zero, where the average is 2.19. 5. Column (5) uses the assigned number of students and its interaction with close peer GPA as instruments for the actual number of students and for its interaction with close peer GPA. 6. Standard errors in parentheses, clustered on the tutorial group level. 7. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.5. Additional outcomes

We investigate whether our findings remain when considering three alternative measures of student performance: course passing rates, credit weighted GPA, and number of credits obtained. To investigate potential selection bias, we also consider the effect of peer ability on the number of courses a student wrote the exam for and student dropout. We analyze these outcomes with our baseline specification in column (4) of Table 3. Course passing rates and an indicator for whether the student wrote the exam are analyzed on the student-course level. The remaining outcomes are analyzed on the student level, where course-cohort fixed effects are replaced with cohort fixed effects.

The results are presented in Table 5. Column (1) reveals that the average grade gains from having better close peers do not occur around the critical threshold of 5.5: the cutoff between a passing and failing grade. However, results from non-linear specifications in Section 6 will show that this zero estimate for passing rates on close peer GPA masks some heterogeneity. Column (2) shows that the positive effects on grades has a cumulative effect on a student's GPA. The estimate indicates that a one standard deviation increase in close peer high school GPA increases a student's credit weighted GPA by roughly 0.04 standard deviations (p -value<0.01). Column (3) is consistent with the finding that close peers do not substantially affect passing rates, and thus have no impact on the total number of credits obtained. Columns (4) and (5) reveal any increase in student performance is not due to the fact that close peers impact the number of courses a student sits the final exam for. Column (6) also shows that close peer GPA does not change the probability of a student dropping out altogether. The latter three columns rule out selection bias. The estimate for distant peer GPA is zero throughout.

5.6. Nature of social interactions

Our baseline results indicate that better classroom peers have small positive implications for a student's performance via peer-to-peer social interactions. What is the nature of these social interactions? A possible answer to this question allows us to speak to the specific examples of the peer effect mechanisms listed within the social interaction category, such as peer-to-peer teaching and effects on study habits. In order to try and answer this question, we will draw on the course evaluations. These results have to be interpreted as suggestive; though unrelated to a student's close and distant peer GPA, the response rate for these self-reported measures is relatively low at 30 percent (see Appendix Table A.8).

Table 5
Peer Effects on Other Outcomes for Student Performance.

	Pass (1) or Fail (0)	GPA Weighted by Credits	Number of Credits	Number of Courses	Followed the Course? Balanced Panel	Dropout
	(1)	(2)	(3)	(4)	(5)	(6)
Close Peer GPA	0.0048 (0.0043)	0.0445*** (0.0147)	0.4530 (0.3462)	0.0284 (0.0483)	0.0028 (0.0048)	0.0004 (0.0080)
Distant Peer GPA	0.0054 (0.0049)	-0.0106 (0.0226)	0.4048 (0.4392)	-0.0140 (0.0585)	-0.0014 (0.0058)	0.0114 (0.0096)
Own GPA	0.1079*** (0.0048)	0.4966*** (0.0177)	8.4899*** (0.3812)	0.4839*** (0.0454)	0.0484*** (0.0045)	-0.0619*** (0.0067)
Observations	18,736	2,218	2,218	2,218	22,180	2,218
Adjusted R ²	0.204	0.303	0.243	0.062	0.033	0.037
Binary Outcome	Yes	No	No	No	Yes	Yes

Notes: 1. All regressions include (course-) cohort fixed effects and controls; student number, a female dummy, age, distance to university, and a non-western and western immigrant dummy. 2. Peer GPA refers to the leave-out mean of high school GPA for the close peer group and to the mean for the distant peer group. Own GPA refers to own high school GPA. All GPA measures are standardized. 3. Column (1) is estimated on the student-course level, where the number of observations is identical to Table 3. Column (5) creates a balanced panel on the student-course level, where the outcome variable takes the value one if a student wrote the final exam for that course and zero otherwise. Column (2), (3), (4) and (6) are estimated on the student level. 4. Column (2) has first-year credit weighted GPA as outcome variable and is based on the number of courses that the student took. Column (3), (4) and (5) refer to the number of credits obtained or the number of courses a student wrote the final exam for. Dropout in column (6) is one if a student did not write an exam in the last block of the first year and zero otherwise. Credit weighted GPA in column (2) is standardized, all other outcomes are unstandardized. Number of credits range from 1 to 60. Number of courses range from 1 to 10. 5. Across the six cohorts there are 78 students (3.4%) who confirmed their registration on the first day but for whom we observe no valid grade. These students dropped out before the first exam week. As we cannot calculate a GPA for them, these students are dropped from this analysis. Results do not change when we include these students. 6. Column (1), (5) and (6) have a binary outcome variable; all columns are estimated with OLS. 7. Standard errors in parentheses, clustered on the tutorial group level. 8. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We start by analyzing students' reported time use. The first three columns of Table 6 repeat the baseline specification to investigate whether close peer high school GPA has an impact upon a student's tutorial attendance (column (1)), lecture attendance (column (2)), and total study time (column (3)). The estimates reveal that a student with better close peers is less likely to attend lectures, while tutorial attendance and reported total study time are not impacted.¹⁹ The estimate in column (2) suggests that a one standard deviation increase in close peer GPA decreases the probability to attend lectures by 1.7 percentage points. Due to the rough (binary) nature of the question, however, we are inclined to interpret only its sign and significance (p -value = 0.03).²⁰ A decrease in lecture attendance, together with no impact upon tutorial attendance and total study time (tutorials + lectures + self study), would suggest that students with better close peers substitute lecture attendance for additional self study.

To further investigate this interpretation, we analyze the impact of close peer high school GPA on the perceived quality of the lecturer and on the perceived usefulness of lectures. Our idea is that students with better close peers might have experienced a lower return to lecture attendance (relative to self study), which suggests that they might have evaluated the lecturer and lectures less positively. The estimates in column (4) and (5) of Table 6 are consistent with this idea; having better close peers decreases the perceived quality of the lecturer and usefulness of the lectures. This further reinforces the interpretation that students might have substituted lecture attendance for additional self study.

Does this potential increase in self study involve close peers studying together outside of the classroom, or do beneficial peer-to-peer interactions only take place during the tutorials, after which individual self study takes place? If the latter is the case, we would expect students' perception of the quality of their TA and the usefulness of tutorials to increase when having better close peers. Column (6) and (7) show, however, that close peer high school GPA is unrelated to students' perceptions of the quality of the TA and the usefulness of the tutorials.²¹ Taken together, while only suggestive, these results suggest that the dependence of peer effects on peer-to-peer social interaction occurs because such effects stem from collaborative self study outside of the classroom. This is consistent with the literature showing sizable behavioral influences from peers (Sacerdote, 2014).

¹⁹ Recall that students are required to attend 70 percent of the tutorials per course, so the possibility for peers to have an impact upon tutorial attendance is limited.

²⁰ For this question students are asked only about the extensive margin of their lecture attendance: "Have you attended lectures?". Even students who attended a few lectures may answer this question with yes (1) instead of no (0). As such, it may well be that these results understate the true reduction in lecture attendance.

²¹ Appendix Table A.8 reveals no effect of students' close peers on the remaining questions regarding their perceptions of the course.

Table 6
Peer Effects on Time Use and Additional Outcomes using Course Evaluations.

	Time Use ¹			Additional Outcomes			
	Tutorial Attendance (1)	Lecture Attendance (2)	Total Study Time (3)	Lecturer Quality (4)	Usefulness Lectures (5)	TA Quality (6)	Usefulness Tutorials (7)
Close Peer GPA	-0.0019 (0.0123)	-0.0170** (0.0077)	-0.1770 (0.1918)	-0.0536*** (0.0201)	-0.0601** (0.0288)	-0.0325 (0.0234)	-0.0088 (0.0285)
Distant Peer GPA	-0.0122 (0.0143)	-0.0047 (0.0086)	0.1717 (0.1899)	-0.0498** (0.0206)	-0.0125 (0.0319)	-0.0110 (0.0222)	0.0044 (0.0302)
Own GPA	0.0370*** (0.0090)	-0.0116 (0.0092)	-0.3703** (0.1537)	0.0267 (0.0212)	0.0218 (0.0221)	0.0002 (0.0192)	-0.0116 (0.0304)
Observations	18,445	4,361	4,361	3,560	2,178	3,560	2,178
Adjusted R ²	0.123	0.137	0.278	0.250	0.252	0.078	0.126
Binary Outcome	No	Yes	No	No	No	No	No

Notes: 1. All regressions include course-cohort fixed effects and controls; student number, a female dummy, age, distance to university, and a non-western and western immigrant dummy. 2. Peer GPA refers to the leave-out mean of high school GPA for the close peer group and to the mean for the distant peer group. Own GPA refers to own high school GPA. All GPA measures are standardized. 3. The dependent variable in column (1) is the standardized percentage of tutorials attended per course, which is drawn from our administrative tutorial attendance data. The remaining outcome variables are drawn from the course evaluations. Column (2) uses the answer to the question "Have you attended lectures?". The dependent variable in column (3) is the answer to the question "Average study time (hours) for this course per week (tutorials+lectures+self study)?" where we used the maximum for the interval to convert the categories into hours. The dependent variables in column (4) and (6) are the mean of the answers to the questions that evaluate the Lecturer/TA. The dependent variables in column (5) and (7) are the answers to the questions "Were the lectures/tutorials useful?". The dependent variables in column (4) until (7) are standardized. 4. Column (2) has a binary outcome variable; all columns are estimated with OLS. 5. Standard errors in parentheses, clustered on the tutorial group level. 6. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6. Non-linear results

While the simplicity of the linear-in-means model is appealing, it may not always accurately capture how peer effects operate (Sacerdote, 2014). Do our previous results concerning the difference in spillovers from close and distant peers hold when adopting more flexible non-linear models?

Following Carrell et al. (2013) we allow for non-linearity by adopting a two-way interaction model. For every student we first calculate the (leave-out) proportion of low, middle, and high ability students in their close and distant peer groups separately. We then estimate models with interaction terms between a student's own ability type and the fraction of high and low ability peers. For each ability type, these interactions show the impact of increasing the proportion of high or low ability peers by decreasing the proportion of average ability peers. For example, Own Low \times Peer High shows the estimated effect on student performance for low ability students of increasing the proportion of high ability peers by decreasing the proportion of average ability peers, either in the close or distant peer group.

We define low and high ability students to be in the bottom and top quartiles of high school GPA across the six cohorts. The middle 50 percent of students are defined as being of average ability. The results are robust to more restrictive categorizations of high and low ability students (available upon request).²²

Table 7 presents our results. Column (1) and (2) first document that our baseline results from the linear-in-means specification carry over to a model where we use the share of high and low ability students to measure peer ability. Next, column (3) and (4) show the heterogeneity results on first-year grades for the close and distant peer model respectively. The results in column (3) reveal spillovers that are roughly linear in close peer ability, implying that the estimates of the linear-in-means model are insightful. Specifically, the estimates show that the observed close peer spillovers are driven primarily by low and high ability students benefiting from an increase in the share of high ability close peers. Both high and low ability students are negatively affected by increasing the share of low ability close peers, insignificantly so for low ability students. Increasing the share of either high or low ability close peers appears to have no impact on average ability students. Conversely, column (4) shows that the proportion of high and low ability peers in one's distant peer group has no significant effect on grades for any ability type.

Column (5) and (6) replace the outcome variable with a pass-fail indicator for the close and distant peer model respectively. In contrast to the linear-in-means estimates, the results reveal that peers do significantly impact first-year passing rates. Similar to grades, peer effects are roughly linear and seem to originate from close peers only. A comparison of columns (3) and (5) does reveal, however, that the close peer spillovers on passing rates are somewhat smaller.

²² Appendix A.6 repeats the balancing test of Table 2 while replacing average peer GPA as the outcome variable separately with the (leave-out) share of low, average, and high ability peers in the close and distant peer group. We find that student characteristics cannot explain the share of peers by ability type; only two out of the 42 estimated coefficients (γ_1 and γ_2) are significant, and the characteristics are always jointly insignificant.

Table 7
Heterogeneity by High School GPA of Peer Effects.

	Grades (Standardized)				Pass (1) or Fail (0)	
	Close (1)	Distant (2)	Close (3)	Distant (4)	Close (5)	Distant (6)
Share Peer High	0.1675*	0.0167				
	(0.0912)	(0.1100)				
Share Peer Low	-0.0473	0.0091				
	(0.1053)	(0.1102)				
Own High × Peer High			0.3679**	-0.0030	0.1653***	0.0929
			(0.1423)	(0.1683)	(0.0563)	(0.0635)
Own High × Peer Low			-0.3193**	0.0901	-0.0304	0.0476
			(0.1498)	(0.1517)	(0.0523)	(0.0587)
Own Avg × Peer High			-0.0389	-0.0621	-0.0096	-0.0106
			(0.1095)	(0.1245)	(0.0444)	(0.0518)
Own Avg × Peer Low			0.1092	0.0371	0.0904*	0.0484
			(0.1204)	(0.1268)	(0.0477)	(0.0545)
Own Low × Peer High			0.3354**	0.2191	0.0848	0.0691
			(0.1498)	(0.1583)	(0.0657)	(0.0702)
Own Low × Peer Low			-0.1311	-0.1661	-0.1038	-0.1462*
			(0.2165)	(0.1985)	(0.0854)	(0.0780)
Observations	18,736	18,736	18,736	18,736	18,736	18,736
Adjusted R ²	0.325	0.324	0.326	0.325	0.205	0.205
F-test Peer Measures	2.72	0.01	3.45	0.60	2.70	1.61
p-value	0.071	0.988	0.004	0.731	0.019	0.154
Binary Outcome	No	No	No	No	Yes	Yes

Notes: 1. All regressions include course-cohort fixed effects, controls; student number, a female dummy, age, distance to university, and a non-western and western immigrant dummy, and own high school GPA. 2. Students are classified into dummies that refer to the bottom 25 percent (low), middle 25 to 75 percent (average), and top 25 percent (high) of high school GPA. The peer measures are the (leave-out) shares of students in the (close) distant peer group belonging to each category. The shares are unstandardized. 3. Odd columns include the leave-out shares for the close peer group and even columns the shares for the distant peer group. 4. The F-test, and corresponding *p*-value, refer to a test for the joint significance of the peer measures included in the model. 5. All columns are estimated with OLS. Column (5) and (6) have a binary outcome variable, estimates using Probit for these two columns can be found in Appendix Table A.7. 6. Standard errors in parentheses, clustered on the tutorial group level. 7. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Column (6) documents that one distant peer measure (Own Low × Peer Low) is statistically significant at the 10%-level. Therefore we conduct two additional analyses to investigate the robustness of our conclusions. First, we perform an F-test on the joint significance of the six peer measures in the models of Table 7. The results are reported in the second-to-last row and confirm that only close peers have a statistically significant impact on passing rates. Second, we replace the linear probability model with a Probit. Appendix A.7 reveals that the sign and significance of the Probit estimates are identical to the peer effects estimates on grades in column (3) and (4) of Table 7. We conclude that our previous results indeed carry over to non-linear models.

The coefficient for Own High × Peer High in column (3) reveals that increasing the share of high ability students by 25 percent, the equivalent of replacing 3 out of 12 average ability students with 3 high ability students, increases the grade of a high GPA student by almost 0.1 standard deviation. When comparing this effect size with other treatments known to have an impact on student performance in higher education, we see that it is roughly twice the size of having a same-sex instructor (Hoffmann and Oreopoulos, 2009), resembles the effect of increasing professor quality by one standard deviation (Carrell and West, 2010), and is similar to the impact of a temporary restriction of legal cannabis access (Marie and Zölitz, 2017). As a further comparison, 0.1 standard deviation corresponds to approximately half of the math gender gap in the fifth grade in the U.S. (Fryer Jr and Levitt, 2010).

7. Close is close, distant is distant

In this section we will analyze first-year tutorial attendance and second-year tutorial and working group registration to investigate whether a student experienced a different degree of social interaction with her close peers than her distant peers. Across all outcomes we find that the close peer group meetings were successful; students experienced a higher degree of social interaction with their close peers (*i.e.*, $1 > s \geq 0$). As such, this section provides independent support for the interpretation that classroom peer effects rely on peer-to-peer social interaction.

7.1. First-Year tutorial attendance

The first analysis we perform exploits our tutorial attendance data on the session level. We ask to what degree a student's attendance at a particular tutorial session is related to the mean attendance of her close and distant peers at that session. Let $Attendance_{isgct}$ be a binary variable taking the value one if student i attended tutorial session s , in group g , for course c , of cohort t . We run the following regression:

$$Attendance_{isgct} = \delta_0 + \delta_1 \overline{Att\ Close}_{(-i)sgc} + \delta_2 \overline{Att\ Distant}_{sgc} + G_{gct} + \delta_3 X_i + \epsilon_{isgct}, \quad (3)$$

where $\overline{Att\ Close}_{(-i)sgc}$ and $\overline{Att\ Distant}_{sgc}$ are the (unstandardized) proportions of attendance at session s of course c from the close and distant peer group.

Eq. 3 regresses attendance on its own group leave-out mean. If one is trying to detect causal peer effects this model would suffer from the reflection problem. The goal of this analysis, however, is to simply estimate the correlation between a student's attendance and the mean attendance of her close and distant peers, as captured by δ_1 and δ_2 respectively. If $\delta_1 > \delta_2$, then a student is more likely to attend a tutorial session together with her close peers than with her distant peers. This would imply an unambiguously higher degree of exposure, and hence social interaction, within tutorial sessions between a student and her close peers than between a student and her distant peers. This is our preferred interpretation of Eq. 3, and does not require a causal reading.

Because of possible session-level attendance shocks, such as bad weather, δ_1 and δ_2 may not reflect deliberate attendance coordination between a student and her peers. We note that under the assumption that common shocks will inflate δ_1 and δ_2 by the same additive constant, a comparison of the magnitude of the two coefficients filters out these common shocks. To capture the common shocks that are present across all tutorial sessions of that course, such as a good TA, Eq. 3 also includes course-tutorial group fixed effects (G_{gct}). To estimate δ_1 and δ_2 , we thus rely on the variation of attendance within a tutorial group across sessions; recall there are fourteen (seven) tutorial sessions for large (small) courses. The remaining control variables (X_i) are identical to the baseline regressions. Similarly, the standard errors are clustered on the tutorial group level.

The close peer group meetings are most salient at the start of the first year, during the first five weeks. To account for this, Eq. 3 is estimated separately for each block. The estimates for δ_1 and δ_2 are presented visually in the top graph of Fig. 3. The full regression results are presented in Appendix A.9. The table also presents p -values of tests for the equality of coefficients within and across blocks.

Before we discuss the results, recall that students are required to attend at least 70 percent of the tutorials per course across the whole year. This requirement makes the scope for detecting differences between the estimates for close and distant peers limited.

From Fig. 3 we identify two main results. First, a student's attendance is more strongly related to the attendance of her close peers than to the attendance of her distant peers. For example, the point estimates for block 2 imply that a 10 percentage point increase in the mean attendance of a student's close (distant) peers at a particular session is associated with a 4.09 (2.33) percentage point increase in a student's probability to attend that session. The difference between these coefficients is statistically significant (p -value = 0.078). For block 1, 4, and 5 the estimate for close peers is also higher, but the p -values for the equality of the coefficients are slightly above 10%.

Secondly, the estimate for close peers shows a large drop after the second block. This drop is statistically significant (p -value = 0.028) and stands in stark contrast to all other smaller and insignificant changes in the estimates across blocks. In particular, the drop after the second block is not visible between students and their distant peers.

These two results suggest that while students experienced a larger degree of social interaction with their close peers than with their distant peers across the whole year, this difference declined after block 2. The smaller estimate for close peers from block 3 onward is consistent with the close peer meetings being most salient at the start of the year. A plausible explanation for the reduction in social ties between close peers is the two-week Christmas break between block 2 and 3, indicated by the dashed vertical line in Fig. 3. This is the only holiday students experience throughout the year.

Assumption (2) of the decomposition in Section 4.1 stipulates that the close peer meetings only had an impact on classroom spillovers through stimulating social interactions. According to this assumption the decline in social interaction between close peers should thus be accompanied by a similar decline in peer effects on student performance. To study this, we repeat the baseline specification of first-year grades on peer GPA per block. The estimates are presented visually in the bottom graph of Fig. 3. The full regression results can be found in Appendix Table A.10. The estimate for close peer GPA is largest in block 1 and 2, then slightly drops in block 3, and becomes a zero in block 4 and 5. The estimate for distant peer GPA is zero throughout. We investigate several potential explanations for the diminishing ability peer effects in the Online Appendix, such as differences in course types across blocks, direct effects of dropout, and measurement error in peer ability due to dropout. We do not find support for any of them. Though we cannot categorically link the decline in ability peer effects across blocks to the decline in social interaction, this pattern does support the validity of our decomposition, and assumption (2) in particular.

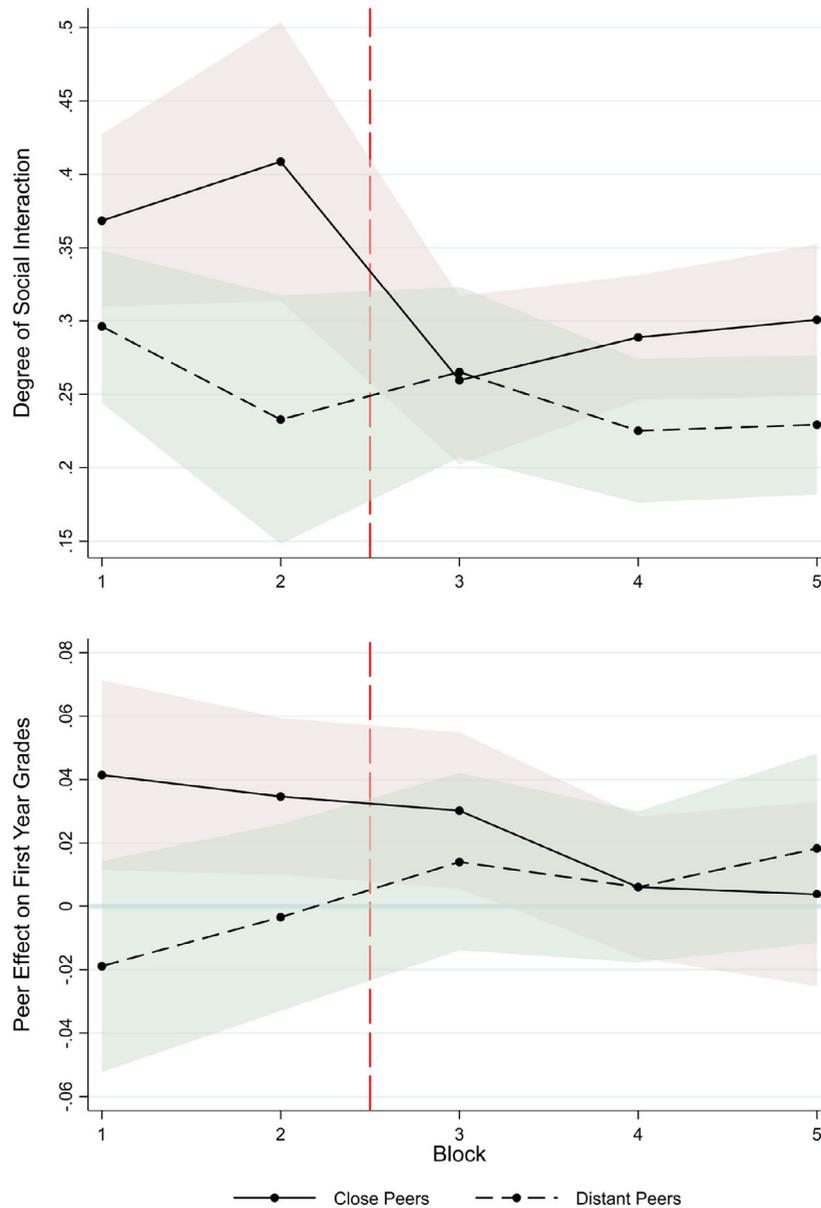


Fig. 3. Diminishing Social Interaction (top graph) and Diminishing Peer Effects in Grades (bottom graph). Top graph shows the point estimates of close and distant peers' average tutorial attendance on a student's attendance per block and the corresponding 90% confidence interval. The precise estimates can be found in Appendix Table A.9. Bottom graph shows the point estimates of close and distant peer effects on first year grades per block and the corresponding 90% confidence interval. The precise estimates can be found in Table. The vertical dashed line indicates the timing of the two-week Christmas break that occurs during the students' first year.

7.2. Second-Year tutorial registration

For the eight second-year courses situated in block 1 to 4, second-year students have to register for a tutorial group a few weeks before the start of the block. Students can choose between roughly fourteen tutorial groups per course. Two students registering for the same two tutorial groups across both courses in the block indicates a strong degree of social interaction.

To analyze co-registration, we adopt the strategy of [Marmaros and Sacerdote \(2006\)](#) and first form all possible pairs of students who are observed to take both courses in block b in the second year of cohort t . Given N_{bt} students this procedure

Table 8
Second Year Tutorial and Working Group Choice.

	Same Tutorial in Second Year?				Same Working Group in Second Year?
	Yes (1) or No (0)				
	Block 1 (1)	Block 2 (2)	Block 3 (3)	Block 4 (4)	Block 5 (5)
Close Peers	0.0630*** (0.0082)	0.0667*** (0.0077)	0.0476*** (0.0054)	0.0641*** (0.0071)	0.0726*** (0.0043)
Distant Peers	0.0402*** (0.0068)	0.0419*** (0.0064)	0.0378*** (0.0060)	0.0379*** (0.0072)	0.0425*** (0.0045)
Both Native Dutch	0.0065*** (0.0019)	0.0070*** (0.0017)	0.0038** (0.0015)	0.0054*** (0.0018)	0.0038*** (0.0005)
Both Non-Western Immigrant	0.0079** (0.0033)	0.0059* (0.0033)	0.0078** (0.0034)	0.0153*** (0.0055)	0.0141*** (0.0019)
Both Western Immigrant	0.0076 (0.0074)	-0.0013 (0.0056)	-0.0061 (0.0048)	0.0065 (0.0101)	0.0049 (0.0041)
Both Female	0.0266*** (0.0051)	0.0275*** (0.0030)	0.0179*** (0.0021)	0.0090*** (0.0026)	0.0244*** (0.0014)
Both Male	0.0031 (0.0021)	0.0015 (0.0019)	0.0025 (0.0017)	0.0052*** (0.0020)	0.0066*** (0.0005)
Same High School	0.1039*** (0.0151)	0.1460*** (0.0147)	0.1158*** (0.0153)	0.1220*** (0.0151)	0.1498*** (0.0126)
Both High GPA	0.0017 (0.0015)	0.0013 (0.0023)	0.0027 (0.0017)	0.0031 (0.0028)	0.0018* (0.0009)
Both Average GPA	-0.0009 (0.0011)	0.0014 (0.0014)	0.0024** (0.0012)	-0.0011 (0.0020)	0.0000 (0.0005)
Both Low GPA	-0.0039 (0.0028)	-0.0037 (0.0030)	-0.0030 (0.0034)	0.0024 (0.0039)	0.0028 (0.0019)
Constant (Without Cohort Fixed Effects)	0.0097*** (0.0020)	0.0122*** (0.0022)	0.0111*** (0.0019)	0.0135*** (0.0018)	-0.0033*** (0.0004)
Observations	133,615	119,937	123,429	98,138	158,302
Adjusted R ²	0.016	0.017	0.014	0.012	0.043
p-value Close = Distant	0.007	0.002	0.071	0.001	0.000

Notes: 1. All regressions include cohort fixed effects. All other explanatory variables that we include are shown in the table. 2. The unit of analysis is a student-pair. The outcome variable is one if both students registered for the same 2 tutorials in a particular block (columns (1)-(4)), or registered for the same working group (column (5)). The explanatory variables are one if both students in the pair share the given characteristic. 3. The outcome variables are binary, the regressions are estimated with OLS. Results using Probit are similar. 4. Constant (Without Cohort Fixed Effects) refers to the constant of an identical regression in which cohort fixed effects have been excluded. Standard error reported in parentheses. 5. The p-value "Close = Distant" tests the equality of the Close Peers and Distant Peers coefficients within a block. 6. Standard errors in parentheses. Following [Cameron et al. \(2011\)](#) we adopt two-way clustering based on the first-year tutorial groups of both students in the pair. 7. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

generates $(N_{bt} \times N_{bt} - 1)/2$ pairings of students.²³ Let $SecondYearTutorial_{ijbt}$ be an indicator variable taking the value of one if student i and j registered to the same two groups for both courses within block b , and zero otherwise. We define a similar set of indicator variables for student characteristics, taking the value of one if students i and j share that particular characteristic and zero otherwise. We then run the following regression per block:

$$SecondYearTutorial_{ijbt} = \pi_0 + \pi_1 SharedCharacteristic_{ij} + C_t + \epsilon_{ijbt}. \quad (4)$$

Coefficient π_1 captures the change in the probability of two students registering for the same two groups in a block if they share that particular characteristic. Eq. 4 includes cohort fixed effects (C_t), but, as the unit of observation is a student pair, it does not include other control variables. To account for the fact that each observation is a student pair, we adopt a two-way clustering standard error procedure and cluster on the first-year tutorial groups of both students in the pair ([Cameron et al., 2011](#)).

The two main characteristics of interest are Close Peers and Distant Peers, which take the value one if students i and j are each others close and distant peers, respectively. We also include various other shared characteristics: immigrant status, gender, former bonds based on a student's high school, and ability (measured by high school GPA). Homophily, observed in virtually every social network, leads people to interact with those who are similar to them. This means that positive estimates for the shared characteristic coefficients serve to validate co-registration as a measure for social interaction.

Column (1) to (4) of [Table 8](#) report the results for block 1 to 4. The last row reports the constant from an identical regression without controlling for cohort fixed effects. The estimates for the shared characteristics may be compared to this constant, which reflects the probability of two students, who do not share any of the included characteristics, registering together for both groups within the block.

²³ This is done by crossing the relevant list of student numbers with itself, removing all duplicate pairs (i, i) , and keeping only one instance of the same pairing (i, j) and (j, i) .

The results reveal two main patterns. First, a comparison of the Close Peers and Distant Peers coefficients reveals that, across blocks, the former is roughly 1.5 to 1.7 times larger than the latter. These differences are statistically significant at the 1%-level. The Close Peers coefficient in block 1 is 0.063, which indicates that sharing a close peer group increases the probability of registering for the same two groups by 6.3 percentage points. Two students with no shared characteristics have a 0.097 percent probability to co-register; 1 out of 100. This becomes roughly 1 out of 14 when the students originate from the same close peer group.

The Distant Peers estimates are also positive and statistically significant. Thus, while distant peers are less important than close peers, a student is more likely to co-register with her distant peers than with someone in a different first-year tutorial group altogether. We view this result as reassuring, as it suggests that the close peer group meetings did not unnaturally alter the classroom environment and lead students to stigmatize their distant peers. This further supports assumption (2) of the decomposition; the close peer meetings only created a set of peers within the classroom with whom social interaction was higher.

The second main pattern is that students tend to co-register with those whom they share characteristics. We find that two students who are both native Dutch, share an immigrant background, have the same gender, or shared the same high school are significantly more likely to appear in the same two tutorial groups. The importance of shared characteristics for co-registration is consistent with homophily and thereby validates our measure of social interaction.²⁴

Although our finding of homophily-like patterns lends support to the idea that co-registration is a reasonable measure of social interaction, we cannot completely reject an explanation where co-registration is observed due to shared preferences on the exact time at which the second-year tutorials are held.²⁵ To address this, we use an alternative measure for social interaction in column (5) of Table 8.

7.3. Second-Year working group registration

In block 5 of the second year students follow a course called “Research Project”. This course revolves around a research report, which students have to write in groups of three to four students. At the start of the course, during the first lecture, students are being told to form groups themselves. One student from each group then registers the working group via the university portal. Students within this working group jointly work on their research report during the whole block. The instruction of this course takes place via two lectures and seven question hours.²⁶

To analyze working group formation, we follow the same procedure as above and form all possible pairs of students who are observed to take the course Research Project in their second year. We then re-estimate Eq. 4 while replacing the outcome variable by $WorkingGroup_{ijt}$, which is equal to one if student i and j of cohort t are in the same working group.

Column (5) of Table 8 documents the results. It shows estimates that are very similar to the ones in column (1) to (4) for all shared characteristics. The most notable difference arises when analyzing the estimates on the Close and Distant Peers. In particular, the former is 1.71 times larger than the latter. This is the largest difference across all columns and it is statistically significant with a p -value of 0.000. This suggests that better measures of social interaction may show a starker difference between a student’s close and distant peers.

8. Conclusion

The promise of the peer effects literature is that simply reorganizing students across classes could increase student performance. Despite many well-identified studies on classroom peer effects, the literature has not yet delivered on this promise. A primary reason for this is our inability to understand the channels at work behind the reduced-form estimates.

This paper tries to address this shortcoming. Our context is an economics undergraduate program of a large public university in the Netherlands. First-year students are randomly assigned to a year-long tutorial group and one of two subgroups within their tutorial group. We exploit a university policy that encourages social interaction within, and not between, these subgroups via a series of informal meetings at the start of the first year. These informal meetings formed two social circles within each tutorial group; we show empirically that social interaction is higher between students within subgroups than students between subgroups. Each student can thus divide her tutorial peers into a set of *close* and *distant* peers.

Exploiting this within-classroom random assignment, we find that classroom peer effects on student performance can solely be attributed to a student’s close peers. We find no role for distant peers. Intuitively, as the degree of social interaction is the only difference between a student’s close and distant peers, this result implies that peer-to-peer social interactions drive classroom peer effects, and that an absence of such interaction prevents peer effects from occurring. Supplementary

²⁴ The estimates on shared high school GPA reveal little to no co-registration of students with similar ability. This is in line with the findings of Marmaros and Sacerdote (2006).

²⁵ To this end, it is useful to note that for each course within a block there are approximately two to three tutorial groups (of in total fourteen) taught at identical times. Thus, students with similar preferences regarding tutorial times could still register in different tutorial groups. We do not, however, observe the time of the second-year tutorial groups.

²⁶ The other second-year course in block 5 is called “Accounting”. The set-up of this course is similar to the second-year courses in block 1 to 4; it has tutorials for which students have to register. We can show that analyzing group registration for this single course in block 5 delivers similar results as in column (1) to (4) of Table 8. Note that, however, block 5 does not have a second course with tutorials, which makes it impossible to analyze an outcome variable which is identical to the ones in column (1) to (4).

data suggests that these social interactions involve collaborative studying outside of class. Our non-linear estimates imply that high and low ability students benefit (suffer) from social interactions with high (low) ability close peers. The implications for policies aiming to exploit peer effects are clear: social interactions between assigned peers must be present in order to yield successful and predictable results.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.euroecorev.2021.103763](https://doi.org/10.1016/j.euroecorev.2021.103763)

References

- Angrist, J.D., 2014. The perils of peer effects. *Labour Econ.* 30, 98–108.
- Arcidiacono, P., Nicholson, S., 2005. Peer effects in medical school. *J. Public Econ.* 89 (2–3), 327–350.
- Athey, S., Imbens, G.W., 2017. The state of applied econometrics: causality and policy evaluation. *Journal of Economic Perspectives* 31 (2), 3–32.
- Booij, A.S., Leuven, E., Oosterbeek, H., 2017. Ability peer effects in university: evidence from a randomized experiment. *Rev. Econ. Stud.* 84 (2), 547–578.
- Brady, R.R., Insler, M.A., Rahman, A.S., 2017. Bad company: understanding negative peer effects in college achievement. *Eur. Econ. Rev.* 98, 144–168.
- Brunello, G., De Paola, M., Scoppa, V., 2010. Peer effects in higher education: does the field of study matter? *Econ. Inq.* 48 (3), 621–634.
- Burke, M.A., Sass, T.R., 2013. Classroom peer effects and student achievement. *J. Labor Econ.* 31 (1), 51–82.
- Cameron, A.C., Gelbach, J.B., Miller, D.L., 2011. Robust inference with multiway clustering. *Journal of Business & Economic Statistics* 29 (2), 238–249.
- Carrell, S.E., Fullerton, R.L., West, J.E., 2009. Does your cohort matter? measuring peer effects in college achievement. *J. Labor Econ.* 27 (3), 439–464.
- Carrell, S.E., Sacerdote, B.L., West, J.E., 2013. From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica* 81 (3), 855–882.
- Carrell, S.E., West, J.E., 2010. Does professor quality matter? evidence from random assignment of students to professors. *Journal of Political Economy* 118 (3), 409–432.
- De Giorgi, G., Pellizzari, M., Redaelli, S., 2010. Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics* 2 (2), 241–275.
- Duflo, E., Dupas, P., Kremer, M., 2011. Peer effects, teacher incentives, and the impact of tracking: evidence from a randomized evaluation in Kenya. *American Economic Review* 101 (5), 1739–1774.
- DUO, 2014. Examenmonitor VO 2014. Technical Report. Dienst Uitvoering Onderwijs.
- Feld, J., Zölitz, U., 2017. Understanding peer effects: on the nature, estimation, and channels of peer effects. *J. Labor Econ.* 35 (2), 387–428.
- Foster, G., 2006. It's not your peers, and it's not your friends: some progress toward understanding the educational peer effect mechanism. *J. Public Econ.* 90 (8–9), 1455–1475.
- Fryer Jr, R.G., Levitt, S.D., 2010. An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics* 2 (2), 210–240.
- Garlick, R., 2018. Academic peer effects with different group assignment policies: residential tracking versus random assignment. *American Economic Journal: Applied Economics* 10 (3), 345–369. doi:[10.1257/app.20160626](https://doi.org/10.1257/app.20160626).
- Guryan, J., Kroft, K., Notowidigdo, M.J., 2009. Peer effects in the workplace: evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics* 1 (4), 34–68.
- Hoffmann, F., Oreopoulos, P., 2009. A professor like me the influence of instructor gender on college achievement. *Journal of Human Resources* 44 (2), 479–494.
- Hong, S.C., Lee, J., 2017. Who is sitting next to you? peer effects inside the classroom. *Quant. Econom.* 8 (1), 239–275.
- Hoxby, C., 2000. Peer effects in the classroom: learning from gender and race variation. NBER Working Paper No. 7867.
- Hoxby, C.M., Weingarth, G., 2005. Taking race out of the equation: school reassignment and the structure of peer effects. Mimeo.
- Jones, T.R., Kofoed, M.S., 2020. Do peers influence occupational preferences? evidence from randomly-assigned peer groups at west point. *J. Public Econ.* 184, 104154.
- Lavy, V., Paserman, M.D., Schlosser, A., 2012. Inside the black box of ability peer effects: evidence from variation in the proportion of low achievers in the classroom. *The Economic Journal* 122 (559), 208–237.
- Lavy, V., Schlosser, A., 2011. Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics* 3 (2), 1–33.
- Lu, F., Anderson, M.L., 2014. Peer effects in microenvironments: the benefits of homogeneous classroom groups. *J. Labor Econ.* 33 (1), 91–122.
- 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.
- Manski, C.F., 1993. Identification of endogenous social effects: the reflection problem. *Rev. Econ. Stud.* 60 (3), 531–542.
- Marie, O., Zölitz, U., 2017. High achievers? cannabis access and academic performance. *Rev. Econ. Stud.* 84 (3), 1210–1237.
- Marmaros, D., Sacerdote, B., 2006. How do friendships form? *Q. J. Econ.* 121 (1), 79–119.
- Oosterbeek, H., Van Ewijk, R., 2014. Gender peer effects in university: evidence from a randomized experiment. *Econ. Educ. Rev.* 38, 51–63.
- Sacerdote, B., 2001. Peer effects with random assignment: results for dartmouth roommates. *Q. J. Econ.* 116 (2), 681–704.
- Sacerdote, B., 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*, 3. Elsevier, pp. 249–277.
- Sacerdote, B., 2014. Experimental and quasi-experimental analysis of peer effects: two steps forward? *Annu. Rev. Econ.* 6 (1), 253–272.
- Stinebrickner, R., Stinebrickner, T.R., 2006. What can be learned about peer effects using college roommates? evidence from new survey data and students from disadvantaged backgrounds. *J. Public Econ.* 90 (8–9), 1435–1454.
- Thiemann, P., 2017. The persistent effects of short-term peer groups in higher education. IZA Discussion Paper No. 11024.
- Zimmerman, D.J., 2003. Peer effects in academic outcomes: evidence from a natural experiment. *Rev. Econ. Stat.* 85 (1), 9–23.