Peer effects in university*

Adam S. Booij[†] Edwin Leuven[‡] Hessel Oosterbeek[§]

April 17, 2013

- Unfinished and preliminary, please do not quote -

Abstract

This paper estimates ability peer effects in higher education. We randomly assigned first year undergraduate students in economics to tutorial groups, conditional on their own prior ability, yielding experimental variation in peer's ability with a wide support. The data contradict models that are linear in moments, shares, or quartiles. We also reject that peer effects are homogenous for students with ability below and above the median. Estimates of the preferred model specifications imply that on average low ability students benefit from tracking while leaving their high achieving peers unaffected. These results suggest that the widespread application of tracking observed in secondary education may also be a beneficial strategy at the graduate level.

JEL-codes: I22; I28

Keywords: Peer effects; Higher Education; Field experiment

1 Introduction

Whenever a new cohort of students enters a school or university, the incoming group is divided into smaller units; classes, workgroups, tutorial groups. In the presence of peer effects at the

*This version: January 2013. We gratefully acknowledge valuable comments from seminar participants in Amsterdam.

[†]University of Amsterdam, Tinbergen Institute, and TIER. adam.booij@uva.nl

[‡]University of Oslo. Also affiliated with CEPR and IZA. edwin.leuven@econ.uio.no

[§]University of Amsterdam, TIER, Tinbergen Institute, CESifo, and FLACSO. h.oosterbeek@uva.nl

level of these units (say workgroups), the assignment of students to workgroups affects their achievement gains. When peer effects are linear in means, the assignment only affects the distribution of achievement. When peer effects are non-linear or heterogeneous, the assignment will also affect students' average achievement. Knowledge about the nature of academic peer effects is indispensable to fully exploit the gains that the assignment of students to workgroups potentially provides.

This paper analyzes ability peer effects among first year undergraduate students in economics at the University of Amsterdam who are placed in the same workgroup. Per cohort there are around 15 of such groups that have an average size of 39. The composition of the groups is fixed for the entire first year and more than 60 percent of all teaching hours take place in these groups. In the academic years 2009, 2010, and 2011, we had permission to manipulate the ability composition of the workgroups, and to randomly assign incoming students to these groups. The manipulation ensures that we have the necessary large variation in ability composition across groups. To illustrate, with random assignment of students to groups, mean standardized GPA per group would for 95% of the groups vary between -0.30 and 0.30. The actual variation in our dataset is between -1 and 1.6. Likewise, with random assignment the standard deviation of standardized GPA per group would for 95% of the groups vary between 0.8 and 1.2. The actual variation in our dataset is from less than 0.3 to more than 1.5 (see subsection for details 2.2, specifically Figure 2). The random assignment of students to groups – conditional on their ability – ensures that students who are of the same ability but are assigned to different groups, are comparable.

The large variation in ability composition across groups is a key feature of our study. Two recent papers in the rapidly expanding literature of ability peer effects in education, demonstrate that one should be careful in extrapolating estimates obtained using a modest range of variation. Carrell et al. (2012) look at freshmen at the United States Air Force Academy. Cadets from the incoming cohorts 2001 to 2006 were randomly assigned to squadrons. Using the naturally occurring (non-manipulated), random variation in peer composition, the authors find that low-ability students benefited significantly from being placed with peers who have high SAT verbal scores, while high-ability and middle-ability students are not significantly affected by the ability composition of their squadron. These results are then used to design an experiment in which half of the incoming freshmen in 2007 and 2008 are – conditional on certain demographics – assigned to squadrons in a way that was intended to maximize the academic achievement of low-ability students. To their surprise the authors find that low-ability students that were exposed to the treatment of optimal peers performed worse than low-ability students that were assigned to the control group. They attribute this to the formation of self-selected homogeneous sub-groups within the treated squadrons.

Hurder (2012) re-analyzes the data of primary school children in Kenya that were originally analyzed in Duflo et al. (2011). In the experiment of Duflo et al., 121 schools that otherwise would have one first grade class received funding to open an extra class. For 60 schools (the treatment group) this was done by separating the top half of the class in terms of initial achievement from the bottom half. For the other 61 schools (the control group) assignment to classes was random. Hurder estimates different specifications of peer effect models in moments, shares, and quartiles using the control group (where the variation in peer composition is modest), and then assesses how well these models predict achievement of students in the treatment group (where the variation in peer composition is substantial). All models perform rather poorly in predicting outcomes for peer effects that are outside the domain on which they have been estimated.[1]

Our estimation results reject models that are linear in moments (mean and standard deviation), shares (top and bottom half, or top, middle and bottom one-thirds) or quartiles (median and interquartile range). The results also reject that ability peer effects are the same for students with ability below and above the mean. We use the estimates of the preferred model specifications to estimate the gains that can be achieved for two prototypical assignments: mixing versus tracking. We also study the "optimal" assignment, given the composition of the total inflow as restriction. The results suggest that low ability students would gain on average 0.30 of a standard deviation of realized credits when moving from mixing to perfect tracking, while leaving their high ability peers unaffected. It is worth noting that these gains can be realized at

[1] Other recent papers on ability peer effects in education include Ammermueller and Pischke (2009), Betts and Shkolnik (2000), Carrell et al. (2009), Hanushek et al. (2003), Lavy et al. (2012b), Lavy et al. (2012a) and Lyle (2009).

zero cost.

On mechanisms [XXX].

This paper proceeds as follows. The next section describes the context, the experimental design, and the data. Section 3 presents and discusses the empirical findings. Section C inquires the robustness of our findings. Section 4 assesses different potential mechanisms explaining our findings. Section 5 summarizes and concludes.

2 Context, design and data

2.1 Context

The experiment was conducted in the academic years 2009, 2010, and 2011, among first years students in the three year bachelor program in economics and business at the University of Amsterdam.[2] In the first year all students in economics and business follow exactly the same program.[3] Students can thus not substitute easy for difficult courses.

Teaching during the first year takes place in the form of central lectures for all first year students together, and in workgroup meetings for groups of at most 40 students. In workgroup meetings students typically receive in depth explanation of the material, ask questions, and practice and discuss exercises and assignments. The instructor of a workgroup is in most cases a member of the permanent academic staff. Students are assigned to a specific workgroup before the start of the year and are supposed to stay in the same group for the entire first year. There were 14 workgroups in 2009, 17 in 2010, and again 17 in 2011.[4]

Table A2 in the appendix lists the first year courses together with their scheduling in the year and their study load in terms of total teaching hours, workgroup hours and credit points. This shows that slightly more than 60% of total teaching hours take place in workgroup meetings. We do not claim that the workgrouplevel is the only level of peer group, or that it is the most

[2]Students meeting the admission requirements are automatically accepted for the study without further selection. The main requirement is that students graduated from the highest (pre-university) track in Dutch secondary education.

[3]Only in their second academic year students can choose different courses to specialize either in economics or in business.

[4]In 2009 we drop two groups with late registrations, and in 2010/2011 we drop a group with students that want to pursue the fiscal economics track in the second year. These groups that were not randomly assigned, and are not part of the experiment.

relevant one. Students can – and will – also interact with students from other workgroups, or even from other studies. Or, in the other direction, students can form informal subgroups within their workgroup of students with whom they interact more frequently.[5] The level of workgroups is, however, the level at which the university assigns the cohort of incoming students to smaller units, and is therefore the level for which information about the pattern of peer effects can be exploited to raise achievement.

2.2 Design

To acquire information about the nature of ability peer effects in workgroups, we conducted an experiment that manipulated the ability composition of first year workgroups and that randomly assigned students to these groups. Students have to be assigned to workgroups before the start of the academic year. The information about their ability that we have at our disposal at this stage is their GPA on the final exams in secondary school,[6] measured in three categories. Table 1 shows the distribution of these categories, and the GPA interval they represent, prior to the start of the academic year, conditional on the type of math – either A or B – the student attended in high school.

Table 1. Prior distribution of GPA category given Math B

		Cohort					
		2009		2010		2011	
		Math B		*Math B*		*Math B*	
GPA interval	*GPAcat*	0	1	0	1	0	1
$GPA < 6\frac{1}{2}$	0	0.27	0.33	0.30	0.33	0.35	0.34
$6\frac{1}{2} \leq GPA < 7$	1	0.36	0.39	0.45	0.37	0.43	0.39
$GPA \geq 7$	2	0.36	0.28	0.25	0.30	0.22	0.26
		1.00	1.00	1.00	1.00	1.00	1.00

We refer to the different categories, which contain roughly 32% (GPA below 6.5), 40% (GPA at least 6.5 but below 7), and 28% (GPA at least 7) of the students, as *GPAcat* 0, 1, and 2.

[5] Defining the relevant peer group is not obvious. Some studies explore this issue by defining peer groups at different levels. For example, Sacerdote (2001) examines peer effects of roommates as well as of dormmates.

[6] High school GPA is the average grade over seven (or more) subjects that are graded on a scale from 1 to 10, where 6 means a pass.

Using this division we manipulated the shares of each type in each workgroup, by setting different assignment probabilities for each workgroup conditional on type. We aimed at creating large variation in the ability composition of workgroups by covering the complete probability triangle in Figure 1. Each dot in the triangle resembles a workgroup as a combination of different *GPAcat* shares. Dots in the three corners resemble workgroups that consist of one type of students only. Dots on the line segments resemble workgroups that consist of two types of students, and dots in the triangle resemble workgroups that consist of three types of students.

In 2010 and 2011 we did the conditional random assignment just before the start of the academic year (September), and set the probabilities such that the prior distributions in Table 1 gave groups of equal size. In 2009 we used rolling assignment where students were assigned at the moment of registration (between June – September). Because of a negative correlation between the moment of registration, henceforth *Application Order*, and the *GPA* of new candidates, the higher ability groups were filled more quickly in this procedure and therefore closed sooner. As this may generate a correlation between *Application Order* and peer ability in the assigned group for this cohort, we include this variable in all our regressions.[7] Also, as students with math type B, henceforth *Math B*, were traditionally grouped together, we treat this group separately. The complete list of assignment probabilities is given in Table A1 in the appendix.

The intent of the assignment procedure is to create large variation in peer ability. After students make their final enrollment decision we obtain their exact *GPA* from the student registry. Figure 2 shows that the variation in the ability composition of workgroups that we obtained through our experimental manipulation is substantially larger than the variation that would have occurred naturally. With unconditional random assignment of students to groups, mean standardized GPA per group would, for 95% of the groups, vary between -0.30 and 0.30. The actual variation in our dataset is between -1 and 1.6. Likewise, with unconditional random assignment the standard deviation of standardized GPA per group would, for 95% of the groups, vary between 0.8 and 1.2. The actual variation in our data is from less than 0.3 to more than 1.5. Hence, the variation that we have is larger than the naturally occurring variation that Carrell

[7] Another reason why some groups were filled more rapidly, is that we could not use the true 2009 prior distribution for setting the probabilities, but had to use the previous *GPA* distribution as a proxy, as the true distribution is known only when all potential new entrants have registered in September.

Figure 1. Prior peer group composition

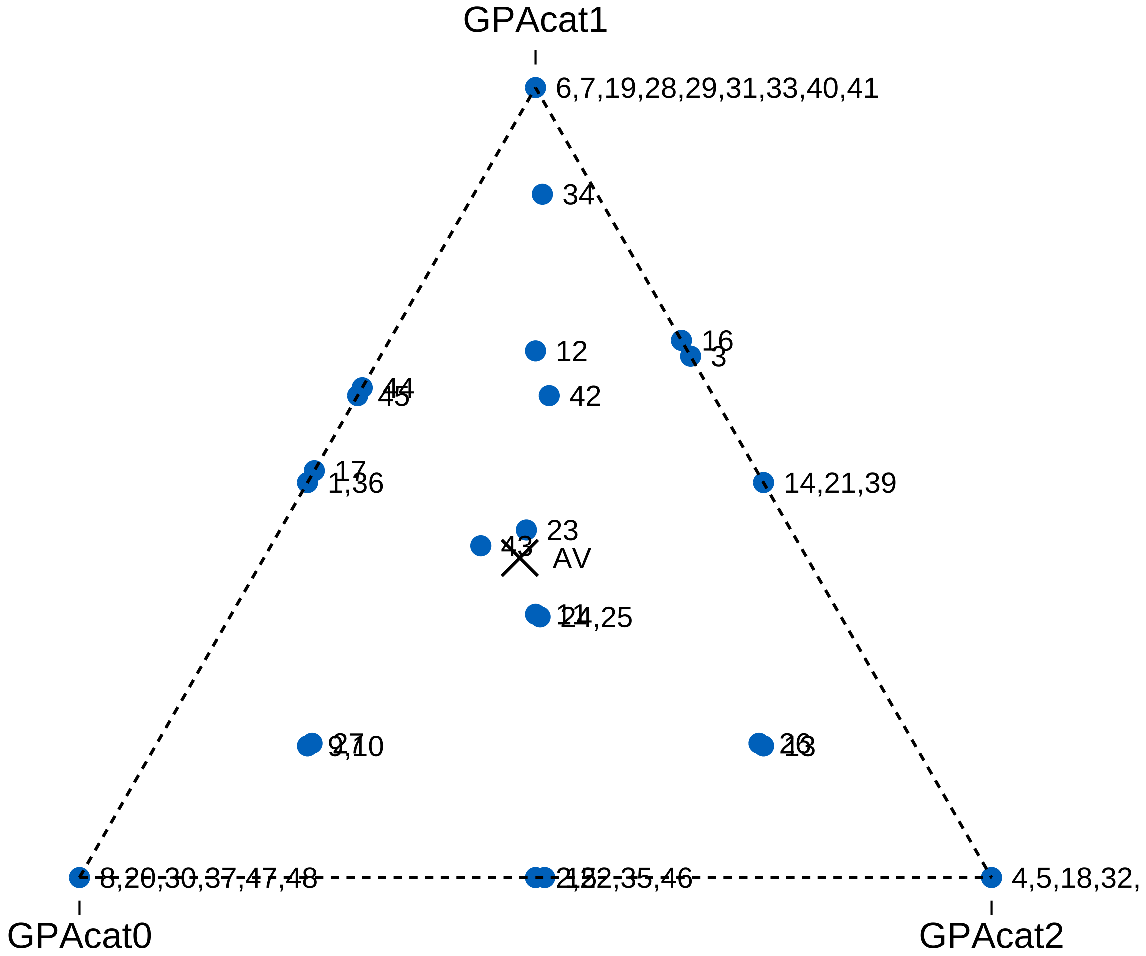
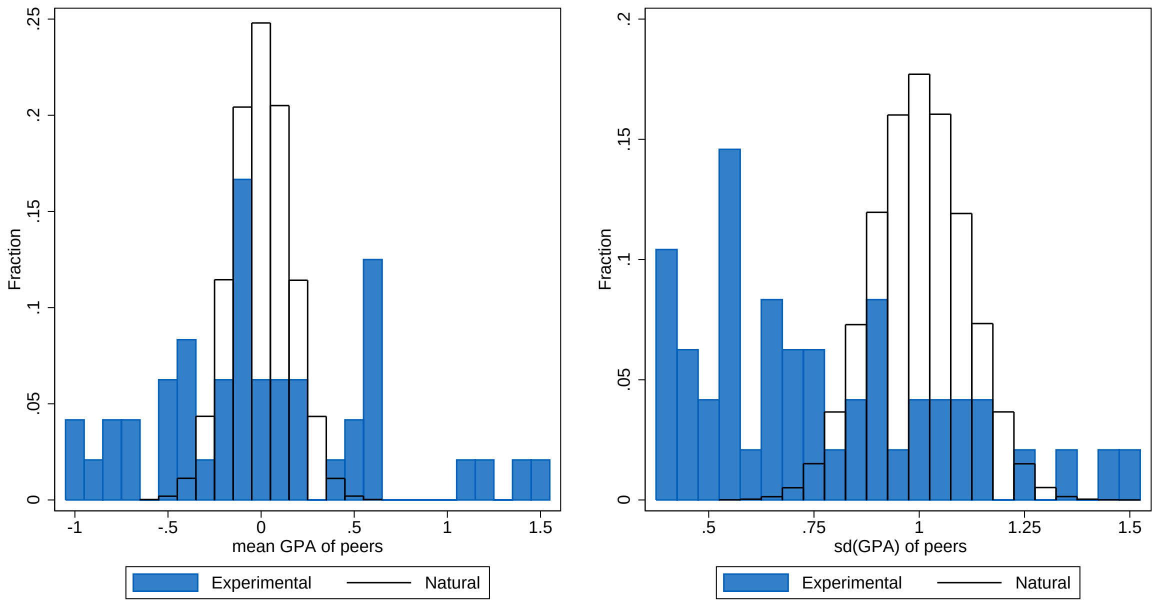


Figure 2. Experimental variation in peer's mean GPA and s.d.



et al. (2012) exploit when they estimate peer effects using the incoming cohorts from 2001 to 2006, and the variation Hurder (2012) exploits when she estimates peer effects using the control group of the experiment of Duflo et al. (2011). The large support will allow us to more precisely detect non-linearities that may give rise to optimal re-allocations.

2.3 Data

Our main data come from the student administration of the department of economics and business of the University of Amsterdam.[8] This source contains information on students' gender, birth date, grades on the final exams in secondary education, the assigned workgroup and their study performance and study status during the first year. Table 2 reports summary statistics, separately for the three cohort. Panel A shows that almost three quarters of the students is male and that the average age at entrance is somewhat above 19 years old. Students who enroll without any delay, would on average enter at the age of 18.5. A substantial share of the students thus enters with a delay of one year or more. These statistics do not vary much across the three cohorts. Students can also enroll in university after studying in a higher vocational school (professional college). The last row of panel A shows that the fraction of students coming through

[8]We also collected additional data through a survey amongst students. We describe (and report about) this data source in Section 4.

Table 2. Summary statistics

		Cohort					
		2009		2010		2011	
A: Background Characteristics	Range	mean	s.d.	mean	s.d.	mean	s.d.
Male	{0,1}	0.73	0.44	0.73	0.44	0.74	0.44
Age	[16.1, 30.1]	19.4	1.56	19.4	1.60	19.4	1.46
Higher Vocational	{0,1}	0.05	0.22	0.04	0.20	0.04	0.19
B: Randomization controls							
raw high school GPA	[5.87, 8.62]	6.71	0.47	6.65	0.46	6.61	0.47
- cat 0: $GPA < 6\frac{1}{2}$	{0,1}	0.33	0.47	0.40	0.49	0.44	0.50
- cat 1: $6\frac{1}{2} \leq GPA < 7$	{0,1}	0.40	0.49	0.35	0.48	0.36	0.48
- cat 2: $GPA \geq 7$	{0,1}	0.26	0.44	0.24	0.43	0.21	0.41
Math B	{0,1}	0.26	0.44	0.37	0.48	0.37	0.48
Application Order	[0,1]	0.47	0.29	0.50	0.29	0.50	0.29
C: Treatment variables							
Mean GPA peers	[-1.03, 1.58]	-0.01	0.57	-0.01	0.54	0.01	0.63
Standard dev. GPA peers	[0.32, 1.52]	0.81	0.29	0.80	0.28	0.74	0.30
D: Outcome Variables							
Credits (raw)	[0,60]	32.1	22.1	34.7	23.3	32.2	24.7
Grade (raw)	[1.00 9.32]	5.26	1.36	5.57	1.42	5.31	1.71
Quit Early	{0,1}	0.11	0.31	0.13	0.34	0.28	0.45
Pass	{0,1}	0.45	0.50	0.54	0.50	0.54	0.50
Number of workgroups		14		17		17	
Number of students		606		668		602	

Note:

this route is small.

Panel B reports summary statistics of students' GPA on the final exams in secondary school. The possible range for students passing the final exams is from 5.87 to 10.[9] The actual range of the GPA of students entering the department of economics and business of the University of Amsterdam is from 5.87 to 8.6. The average GPA is around 6.7 with a standard deviation close to 0.5. Average GPA is slightly (but significantly) lower for the 2010 and 2011-cohorts than the 2009-cohort; the difference between the 2010 and 2011 cohorts is not significant.

Panel C reports summary statistics of the treatment variables. We summarize the ability composition of the peer who are assigned to the same workgroup in in moments (mean and standard deviation) of the standardized GPA. This shows again the large variation in peer composition in our data set.

[9]The lowest collection of grades on seven subjects that still gives a pass is all 6 and a single 5.

Finally, panel D reports summary statistics of the outcome variables. During the first year, students can collect a maximum of 60 credit points if they pass all first year courses. The share that manages to do so is low (21%), and the average number of collected credit points is slightly above 30. This shows that there is quite some scope for improvement; we will not fail to find peer effects because of ceiling effects. The second outcome measure is the grade point average on the exams taken. While grade point average is a common outcome measure, it is informativeness is less when students do not take all exams which may be selective. Only xx% of the students write all first year exams during the first year. The other yy% of the students miss at least one exam, and on average they miss x.x exams. There is no obvious way to correct grade point average for this missing information, which is why this is not our main outcome measure. The third outcome of interest is the fraction of students that quit early (17%), in February. Finally, we look at the first year pass rate, the fraction of students that actually passes the first year (51%) by obtaining more that 45 credits. Students that do not meet this threshold fail the first year, and are not allowed to continue to the second year. The university's main aim is to improve the first year pass rate.

To assess whether the randomization is valid we examine if background characteristics are balanced across workgroups with different treatments. Columns (1) and (2) of Table 3 show results of regressing the different treatment measures on background characteristics, conditional on students' own ability variables.[10] This shows no systematic patterns, as expected. At the same time, columns (3) to (6) show that the background variables are relevant predictors of the outcomes. Male students and older students collect fewer credits and are less likely to pass, and younger students obtain higher grades.

3 Results

Table 4 shows results for specifications in which peers' ability is captured in the mean GPA of workgroup peers and its standard deviation, and where the outcome is measured as the number

[10] After setting the intended shares of the three ability categories for each workgroup, students were given their ability randomly assigned to workgroups. Clearly, a student from the high ability category has a higher chance to be assigned to a workgroup with a high intended share of high ability students. This is why the regression results are conditional on students' own ability.

Table 3. Balancing

	Treatment		Outcomes			
	(1) Mean GPA	(2) SD GPA	(3) Credits	(4) Grade	(5) Quit Early	(6) Pass
Male	-0.01	0.00	-0.19	-0.11	0.04	-0.11
	(0.03)	(0.01)	(0.04)***	(0.04)**	(0.02)**	(0.02)***
Youngest $\frac{1}{3}$	0.01	-0.00	0.09	0.11	0.03	0.02
	(0.03)	(0.01)	(0.05)*	(0.04)***	(0.02)	(0.02)
Oldest $\frac{1}{3}$	-0.00	-0.02	-0.12	-0.05	-0.01	-0.04
	(0.03)	(0.02)	(0.05)**	(0.04)	(0.02)	(0.02)*
Higher Vocational	-0.04	-0.00	0.02	0.12	-0.06	0.03
	(0.05)	(0.03)	(0.12)	(0.11)	(0.04)*	(0.06)
Randomization controls	✓	✓	✓	✓	✓	✓
$\bar{y}$	-0.00	0.78	0.00	0.00	0.17	0.51
$sd(y)$	0.58	0.29	1.00	1.00	0.38	0.50
χ^2-stat coeff$=0$		3.68	14.50	4.09	4.52	11.55
p-value		0.88	0.00	0.01	0.00	0.00
R^2	0.40	0.20	0.26	0.31	0.13	0.23
N	1876	1876	1876	1753	1876	1876

Note: Robust (col. 1-2) / group clustered (col.3-6) standard errors in parentheses. */**/*** denote significance at a 10/5/1% confidence level.

of credit points.

Figure 3 shows the results graphically

3.1 Peer effects on credits

3.2 Non-linearity

3.3 Heterogeneity by ability

3.4 Tracking versus Mixing

3.5 Robustness

3.6 Other outcomes

4 Mechanisms

At the end of both academic years, we carried out a survey among the students in our experiment. The purpose was to gain further insight into what mechanisms took place in the classroom.

Table 4. Results on credits

	First Year Credits										
	(1)	(2)	(3)	(4)	(5)			(6)			
Peer Prior *GPA*	All	All	All	All	Joint	Below	Above	Joint	Bottom	Middle	Top
$\overline{GPA}_{\neg i}$	0.084	0.080	0.11	0.24		0.40	0.016		0.57	0.18	0.034
	(0.035)**	(0.033)**	(0.037)***	(0.076)***		(0.17)**	(0.12)		(0.17)***	(0.17)	(0.16)
$sd\,(GPA_{\neg i}) - 1$			-0.15	-0.21		-0.34	0.019		-0.50	-0.45	0.19
			(0.072)**	(0.099)**		(0.19)*	(0.15)		(0.21)**	(0.29)	(0.16)
$\overline{GPA}^{2}_{\neg i}$				-0.087		0.19	-0.10		0.093	-0.0058	-0.17
				(0.043)**		(0.14)	(0.071)		(0.18)	(0.17)	(0.11)
$(sd\,(GPA_{\neg i}) - 1)^2$				0.062		0.31	0.090		-0.46	-0.44	0.68
				(0.23)		(0.59)	(0.29)		(0.60)	(0.59)	(0.48)
$\overline{GPA}_{\neg i} \times (sd\,(GPA_{\neg i}) - 1)$				0.35		0.11	-0.38		0.81	0.27	-0.78
				(0.21)		(0.54)	(0.34)		(0.62)	(0.61)	(0.51)
Randomization controls	✓	✓	✓	✓		✓			✓		
Controls		✓	✓	✓		✓			✓		
$\bar{y}$			0.00		0.00	-0.38	0.38	0.00	-0.50	-0.01	0.51
$sd\,(y)$			1.00		1.00	0.94	0.91	1.00	0.92	0.96	0.48
R^2	0.24	0.26	0.26	0.26	0.27			0.27			
$N_{cluster}$			48		48			48			
N			1876		1876	938	938	1876	624	626	626
F-tests (p-values)											
cf($\overline{GPA}$) = previous		0.386	0.050	0.045							
cf(Peer variables) = 0	0.019	0.019	0.013	0.005	0.002	0.001	0.552	0.003	0.009	0.550	0.489
cf(NL terms) = 0			0.045	0.005	0.001	0.000	0.495	0.002	0.006	0.418	0.409
cf(Added terms) = 0	0.019	0.019	0.045	0.025							
cf(Peer var.) = homo.					0.034			0.085			
Predicted Tracking Effect											
Difference (Track - Mix)	0.00	0.00	0.084	0.133	0.146	0.298	-0.006	0.108	0.194	0.090	0.041
s.e.(Track - Mix)			(0.041)**	(0.057)**	(0.061)**	(0.120)**	(0.066)	(0.072)	(0.134)	(0.177)	(0.081)
p-value			0.045	0.023	0.020	0.016	0.933	0.138	0.155	0.614	0.613

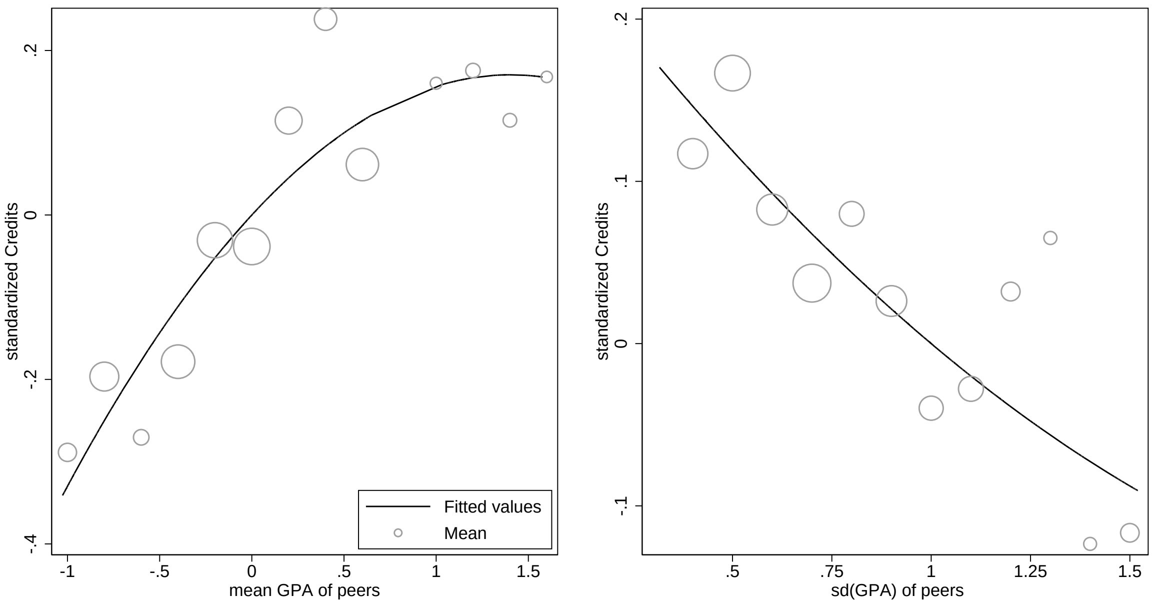
Note: Group clustered standard errors in parenthesis. */**/*** denote significance at a 10/5/1% confidence level.

Table 5. Results on other outcomes.

	Outcome variable											
	Credits			Grade			Quit Early			Pass		
Peer Prior *GPA*	Joint	Below	Above	Joint	Below	Above	Joint	Below	Above	Joint	Below	Above
$\overline{GPA}_{\neg i}$		0.40	0.016		0.35	-0.0018		-0.13	-0.050		0.064	0.026
		(0.17)**	(0.12)		(0.16)**	(0.12)		(0.063)**	(0.038)		(0.076)	(0.055)
$sd\left(GPA_{\neg i}\right)-1$		-0.34	0.019		-0.48	-0.19		0.11	-0.0046		0.012	0.018
		(0.19)*	(0.15)		(0.20)**	(0.19)		(0.072)	(0.050)		(0.090)	(0.080)
$\overline{GPA}^2_{\neg i}$		0.19	-0.10		-0.022	-0.13		0.0015	0.025		0.17	-0.065
		(0.14)	(0.071)		(0.13)	(0.069)*		(0.057)	(0.023)		(0.058)***	(0.035)*
$\left(sd\left(GPA_{\neg i}\right)-1\right)^2$		0.31	0.090		-0.18	-0.51		-0.18	-0.15		0.55	0.082
		(0.59)	(0.29)		(0.62)	(0.33)		(0.20)	(0.10)		(0.24)**	(0.14)
$\overline{GPA}_{\neg i}\times\left(sd\left(GPA_{\neg i}\right)-1\right)$		0.11	-0.38		0.28	-0.43		-0.23	-0.12		-0.25	-0.12
		(0.54)	(0.34)		(0.56)	(0.41)		(0.18)	(0.11)		(0.23)	(0.18)
Randomization controls		✓			✓			✓			✓	
Controls		✓			✓			✓			✓	
$\bar{y}$	0.00	-0.38	0.38	0.00	-0.45	0.43	0.17	0.25	0.10	0.51	0.33	0.69
$sd\left(y\right)$	1.00	0.94	0.91	1.00	0.85	0.94	0.38	0.43	0.29	0.50	0.47	0.46
R^2	0.27			0.33			0.14			0.24		
$N_{cluster}$	48			48			48			48		
N	1876	938	938	1753	855	898	1876	938	938	1876	938	938
F-tests (p-values)												
cf(Peer variables) = 0	0.002	0.001	0.552	0.001	0.008	0.018	0.000	0.016	0.005	0.004	0.001	0.513
cf(NL terms) = 0	0.001	0.000	0.495	0.001	0.009	0.014	0.000	0.008	0.006	0.002	0.001	0.380
cf(Peer var.) = homo.	0.034			0.040			0.462			0.006		
Difference (Track - Mix)	0.146	0.298	-0.006	0.013	0.118	-0.093	-0.107	-0.184	-0.030	0.106	0.219	-0.006
s.e.(Track - Mix)	(0.061)**	(0.120)**	(0.066)	(0.076)	(0.130)	(0.078)	(0.029)***	(0.058)***	(0.025)	(0.034)***	(0.060)***	(0.036)
p-value	0.020	0.016	0.933	0.869	0.369	0.241	0.001	0.003	0.222	0.003	0.001	0.875

Note: Group clustered standard errors in parenthesis. */**/*** denote significance at a 10/5/1% confidence level.

Figure 3. Effect of peer's mean GPA and s.d. on first year credits



5 Summary and discussion

References

Ammermueller, A. and Pischke, J.-S. (2009). Peer effects in european primary schools: Evidence from the progress in international reading literacy study. *Journal of Labor Economics*, 27(3):315–348.

Betts, J. R. and Shkolnik, J. L. (2000). The effects of ability grouping on student achievement and resource allocation in secondary schools. *Economics of Education Review*, 19(1):1–15.

Carrell, S. E., Fullerton, R. L., and West, J. E. (2009). Does your cohort matter? measuring peer effects in college achievement. *Journal of Labor Economics*, 27(3):439–464.

Carrell, S. E., Sacerdote, B. I., and West, J. E. (2012). From natural variation to optimal policy? an unsuccessful experiment in using peer effect estimates to improve student outcomes. Unpublished working paper.

Duflo, E., Dupas, P., and 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.

Hanushek, E. A., Kain, J. F., Markman, J. M., and Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, 18(5):527–544.

Hurder, S. (2012). Evaluating econometric models of peer effects with experimental data. Unpublished working paper.

Lavy, V., Paserman, M. D., and Schlosser, A. (2012a). Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom*. *Economic Journal*, 122(559):208–237.

Lavy, V., Silva, O., and Weinhardt, F. (2012b). The good, the bad, and the average: Evidence on ability peer effects in schools. *Journal of Labor Economics*, 30(2):pp. 367–414.

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.

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *Quarterly Journal of Economics*, 116(2):681–704.

A Assignment probabilities

See table A1 on page 17.

B Overview first year courses

Table A2. Overview of the first year courses in the economics and business program

Course	Term	Total teaching hours	Workgroup hours	Credit points
Financial accounting	1	28	14	5
Organization	1	12	12	5
Orientation fiscal economics	1	6	0	2
Mathematics 1	1/2	56	28	5
Academic skills 1	1/2	28	28	2
Management accounting	2	28	14	4
Microeconomics	2	42	28	7
Organization and management	3	28	14	6
Statistics	3	42	14	5
Mathematics 2	3/4	56	28	4
Academic skills 2	3/4	28	28	3
Finance	4	21	21	5
Macroeconomics	4	42	28	7
Total		417	257	60

C Robustness

C.1 Other peer models

See table A3 on page 18.

C.2 Other potentially confounding peer characetristics

See table A4 on page 19.

C.3 Stability over time

See table A5 on page 20.

Table A1. Group assignment probabilities conditional on GPA category and Math B

	Cohort													
	2009					2010					2011			
	GPAcat					*GPAcat*					*GPAcat*			
Group	0	1	2	*MathB*	Group	0	1	2	*MathB*	Group	0	1	2	*MathB*
1	0.50	0.43	0.00	1	15	0.25	0.00	0.28	1	32	0.00	0.00	0.63	1
2	0.50	0.00	0.60	1	16	0.00	0.31	0.17	1	33	0.00	0.42	0.00	1
3	0.00	0.57	0.40	1	17	0.25	0.23	0.00	1	34	0.03	0.37	0.05	1
4	0.00	0.00	0.25	0	18	0.00	0.00	0.55	1	35	0.24	0.00	0.32	1
5	0.00	0.00	0.25	0	19	0.00	0.46	0.00	1	36	0.24	0.21	0.00	1
6	0.00	0.25	0.00	0	20	0.51	0.00	0.00	1	37	0.49	0.00	0.00	1
7	0.00	0.25	0.00	0	21	0.00	0.10	0.18	0	38	0.00	0.00	0.41	0
8	0.33	0.00	0.00	0	22	0.15	0.00	0.18	0	39	0.00	0.11	0.20	0
9	0.22	0.04	0.04	0	23	0.09	0.09	0.10	0	40	0.00	0.21	0.00	0
10	0.22	0.04	0.04	0	24	0.10	0.07	0.12	0	41	0.00	0.21	0.00	0
11	0.11	0.08	0.08	0	25	0.10	0.07	0.12	0	42	0.05	0.13	0.09	0
12	0.06	0.17	0.04	0	26	0.05	0.03	0.24	0	43	0.09	0.09	0.09	0
13	0.06	0.04	0.17	0	27	0.20	0.03	0.06	0	44	0.10	0.13	0.00	0
14	0.00	0.12	0.12	0	28	0.00	0.20	0.00	0	45	0.10	0.13	0.00	0
					29	0.00	0.20	0.00	0	46	0.13	0.00	0.20	0
					30	0.30	0.00	0.00	0	47	0.26	0.00	0.00	0
					31	0.00	0.20	0.00	0	48	0.26	0.00	0.00	0

Note: Group

Table A3. Results on credits of other peer statistics

	Moment Based				Quantile Based				Share Based		
	(1)				(2)				(3)		
Peer Prior *GPA*	Joint	Above	Below	Peer Prior *GPA*	Joint	Above	Below	Peer Prior *GPA*	Joint	Above	Below
$\overline{GPA}_{\neg i}$		0.40	0.016	$MED_{\neg i}$		0.26	-0.019	$FB_{\neg i} - \frac{1}{3}$		-0.84	-0.078
		(0.17)**	(0.12)			(0.11)**	(0.065)			(0.26)***	(0.20)
$sd\left(GPA_{\neg i}\right) - 1$		-0.34	0.019	$IQR_{\neg i} - 1.35$		-0.16	0.017	$FT_{\neg i} - \frac{1}{3}$		-0.082	0.18
		(0.19)*	(0.15)			(0.080)*	(0.074)			(0.18)	(0.24)
$\overline{GPA}^2_{\neg i}$		0.19	-0.10	$MED^2_{\neg i}$		0.077	-0.054	$\left(FB_{\neg i} - \frac{1}{3}\right)^2$		0.61	0.33
		(0.14)	(0.071)			(0.14)	(0.071)			(0.91)	(0.85)
$\left(sd\left(GPA_{\neg i}\right) - 1\right)^2$		0.31	0.090	$\left(IQR_{\neg i} - 1.35\right)^2$		0.11	0.078	$\left(FT_{\neg i} - \frac{1}{3}\right)^2$		0.24	-0.033
		(0.59)	(0.29)			(0.15)	(0.091)			(1.39)	(0.46)
$\overline{GPA}_{\neg i} \times \left(sd\left(GPA_{\neg i}\right) - 1\right)$		0.11	-0.38	$MED_{\neg i} \times \left(IQR_{\neg i} - 1.35\right)$		-0.050	-0.22	$\left(FB_{\neg i} - \frac{1}{3}\right) \times \left(FT_{\neg i} - \frac{1}{3}\right)$		-1.71	0.82
		(0.54)	(0.34)			(0.20)	(0.11)**			(1.96)	(1.16)
Randomization controls		✓				✓				✓	
Controls		✓				✓				✓	
$\bar{y}$	0.00	-0.38	0.38		0.00	-0.38	0.38		0.00	-0.38	0.38
$sd\left(y\right)$	1.00	0.94	0.91		1.00	0.94	0.91		1.00	0.94	0.91
R^2	0.27				0.27				0.27		
$N_{cluster}$	48				48				48		
N	1876	938	938		1876	938	938		1876	938	938
$AICc$	4832.7				4837.4				4834.9		
F-tests (p-values)											
cf(Peer variables) = 0	0.002	0.001	0.552		0.017	0.008	0.370		0.012	0.005	0.804
cf(NL terms) = 0	0.001	0.000	0.495		0.020	0.012	0.255		0.017	0.003	0.792
cf(Peer var.) = homo.	0.034				0.126				0.082		
Predicted Tracking Effect											
Difference (Track - Mix)	0.146	0.298	-0.006		0.057	0.081	0.033		0.095	0.167	0.022
s.e.(Track - Mix)	(0.061)**	(0.120)**	(0.066)		(0.049)	(0.089)	(0.048)		(0.104)	(0.197)	(0.091)
p-value	0.020	0.016	0.933		0.251	0.371	0.505		0.367	0.400	0.811

Note: Group clustered standard errors in parenthesis. */**/*** denote significance at a 10/5/1% confidence level.

Table A4. Results on credits including other peer characteristics

Peer Prior *GPA*	(1)	(2)	(3)	(4)	(5)
$\overline{GPA}_{\neg i}$	0.24	0.24	0.25	0.17	0.17
	(0.076)***	(0.089)**	(0.077)***	(0.090)*	(0.099)*
$sd\left(GPA_{\neg i}\right)-1$	-0.21	-0.20	-0.24	-0.17	-0.21
	(0.099)**	(0.10)*	(0.10)**	(0.10)	(0.11)*
$\overline{GPA}^2_{\neg i}$	-0.087	-0.087	-0.094	-0.059	-0.070
	(0.043)**	(0.043)**	(0.043)**	(0.052)	(0.053)
$\left(sd\left(GPA_{\neg i}\right)-1\right)^2$	0.062	0.065	0.10	0.077	0.12
	(0.23)	(0.24)	(0.22)	(0.22)	(0.21)
$\overline{GPA}_{\neg i}\times\left(sd\left(GPA_{\neg i}\right)-1\right)$	0.35	0.35	0.38	0.26	0.29
	(0.21)	(0.22)	(0.20)*	(0.20)	(0.20)
$\overline{FBoys}_{\neg i}$		✓			✓
$\overline{Age}_{\neg i}$			✓		✓
$\overline{App.Order}_{\neg i}$				✓	✓
Randomization controls	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓
R^2	0.26	0.26	0.26	0.26	0.26
$N_{cluster}$	48	48	48	48	48
N	1876	1876	1876	1876	1876
F-tests (p-values)					
cf(Peer variables) = equal to (1)		1.000	0.590	0.669	0.337

Note: Group clustered standard errors in parenthesis. */**/*** denote significance at a 10/5/1% confidence level.

Table A5. Results on credits for different years

		Excluded Cohort		
	(1)	(2)	(3)	(4)
Peer Prior *GPA*	All	2009	2010	2011
$\overline{GPA}_{\neg i}$	0.24	0.26	0.28	0.20
	(0.076)***	(0.11)**	(0.11)**	(0.073)**
$sd\,(GPA_{\neg i}) - 1$	-0.21	-0.18	-0.20	-0.24
	(0.099)**	(0.14)	(0.14)	(0.10)**
$\overline{GPA}^{2}_{\neg i}$	-0.087	-0.059	-0.13	-0.13
	(0.043)**	(0.057)	(0.050)**	(0.081)
$(sd\,(GPA_{\neg i}) - 1)^2$	0.062	-0.20	0.25	0.25
	(0.23)	(0.30)	(0.21)	(0.40)
$\overline{GPA}_{\neg i} \times (sd\,(GPA_{\neg i}) - 1)$	0.35	0.53	0.068	0.068
	(0.21)	(0.33)	(0.24)	(0.37)
Randomization controls	✓	✓	✓	✓
Controls	✓	✓	✓	✓
R^2	0.26	0.28	0.25	0.25
$N_{cluster}$	48	34	31	31
N	1876	1270	1208	1274
F-tests (p-values)				
cf(Peer variables) = equal to (1)		0.839	0.303	0.299

Note: Group clustered standard errors in parenthesis. */**/*** denote significance at a 10/5/1% confidence level.