Uncovering Peer Effects in Social and Academic Skills

Román Andrés Zárate

Online Appendix

A Supplementary Material

TABLE A.1: Correlation of Sociability and Social Skills Outcomes

	Big Five Personality Traits					Peers' Perception				Other measures of social skills	Social Skills Index at baseline	Social Skills Index at endline
	Openness	Conscientiousness	Emotional Stability	Extraversion	Agreeableness	Leadership	Friendliness	Popularity	Shyness			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Academic achievement	0.090***	-0.001	0.065***	-0.015	-0.015	0.217***	0.044***	0.105***	-0.066***	0.040**	0.056***	0.040**
	(0.021)	(0.022)	(0.022)	(0.019)	(0.020)	(0.020)	(0.016)	(0.022)	(0.019)	(0.016)	(0.017)	(0.017)
Sociability	0.103***	0.062**	0.032*	0.142***	0.121***	0.230***	0.360***	0.215***	-0.103***	0.092***	0.125***	0.117***
	(0.024)	(0.025)	(0.017)	(0.021)	(0.017)	(0.021)	(0.020)	(0.029)	(0.021)	(0.019)	(0.019)	(0.020)
Social-fit score	0.072***	0.022	0.001	0.063***	0.027	0.136***	0.075***	0.103***	-0.112***	0.027	0.057***	0.042**
	(0.018)	(0.020)	(0.020)	(0.019)	(0.020)	(0.015)	(0.016)	(0.017)	(0.020)	(0.018)	(0.018)	(0.019)
Interview score	0.092***	0.072***	0.072***	0.090***	0.048***	0.069***	0.058***	0.050***	-0.036**	0.066***	0.118***	0.080***
	(0.019)	(0.016)	(0.016)	(0.018)	(0.018)	(0.015)	(0.015)	(0.017)	(0.017)	(0.016)	(0.016)	(0.015)
N	3,106	3,106	3,106	3,106	3,106	3,637	3,637	3,637	3,637	3,654	3,654	3,654

Notes: This table reports standardized estimates of an OLS regression on social skills outcomes of social centrality at baseline and the score in the three tests of the admission process to the COAR Network. All regressions include school-by-grade-by-gender fixed effects. Academic achievement and social centrality are measured at baseline. Centrality at baseline is measured by the eigenvector centrality of an aggregate social network of dorm preferences, friendships, study, and social partnerships. Eigenvector centrality is a measure of the influence of a student in the network. Academic achievement at baseline is the score on the admission test to the COAR Network, which evaluates the applicants in math and reading comprehension. In columns 1 to 5, the dependent variables are personality traits from the Big Five. In columns 6 to 9, the dependent variables are the number of peers who perceive the student as part of the top 5 of leadership, friendliness, popularity, and shyness. In column 10, the dependent variable is an index excluding social network outcomes, personality traits, and peers' perceptions. In columns 11 and 12, the dependent variable is a social skills index that excludes social network outcomes. Column 11 presents the correlations on this index at baseline and column 12 at endline. All indexes are constructed using Principal Component Analysis (PCA) on all the variables that measure social skills (see Appendix D for details). *** p-value<0.01, ** p-value<0.05, * p-value<0.1.

Table A.2: Balance Tests for the More Central Peers Treatment

Variable	All Students		Less Central at Baseline		More Central at Baseline	
	Control mean	Difference	Control mean	Difference	Control mean	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
Admission test	-0.016	0.002	-0.064	-0.024	0.058	0.028
		(0.019)		(0.027)		(0.026)
Interview score	-0.006	0.017	-0.006	0.041	-0.006	-0.006
		(0.030)		(0.042)		(0.042)
Social-fit score	-0.024	0.045	-0.028	0.027	-0.019	0.062
		(0.031)		(0.042)		(0.044)
Female (%)	0.591	0.000	0.589	0.000	0.594	0.000
		(0.000)		(0.000)		(0.000)
Poor (%)	0.431	0.014	0.460	0.026	0.389	0.003
		(0.014)		(0.021)		(0.019)
Rural household (%)	0.284	-0.018	0.314	-0.024	0.241	-0.013
		(0.014)		(0.020)		(0.019)
Subsidized health insurance	0.508	0.008	0.552	-0.011	0.443	0.027
		(0.016)		(0.022)		(0.022)
Math scores	-0.043	0.027	-0.141	0.048	0.103	0.007
		(0.023)		(0.032)		(0.031)
Reading scores	-0.029	-0.001	-0.128	0.038	0.120	-0.040
		(0.021)		(0.031)		(0.028)
Social skills	-0.108	-0.000	-0.522	0.023	0.515	-0.024
		(0.022)		(0.024)		(0.037)
Degree friends	7.331	0.230	5.609	0.163	9.918	0.295
		(0.130)		(0.113)		(0.234)
Centrality friends	-0.110	0.021	-0.521	0.032	0.505	0.011
		(0.025)		(0.023)		(0.044)
Degree study	4.560	-0.007	3.741	-0.060	5.789	0.046
		(0.069)		(0.081)		(0.111)
Centrality study	-0.071	-0.006	-0.350	-0.056	0.340	0.043
		(0.027)		(0.028)		(0.046)
Degree all	10.475	0.067	8.261	-0.060	13.802	0.193
		(0.135)		(0.127)		(0.238)
Centrality all	-0.152	0.023	-0.705	-0.010	0.679	0.056
		(0.018)		(0.016)		(0.031)
Reading the mind in the eyes	20.521	-0.060	20.224	0.184	20.960	-0.304
		(0.130)		(0.186)		(0.180)
Peers' perception leadership	2.499	-0.185	1.586	-0.111	3.874	-0.259
		(0.150)		(0.132)		(0.271)
Peers' perception friendliness	2.550	-0.008	1.836	0.195	3.625	-0.214
		(0.080)		(0.083)		(0.135)
Peers' perception popularity	2.201	0.108	1.465	0.127	3.310	0.089
		(0.159)		(0.142)		(0.285)
Peers' perception shyness	2.083	0.025	2.561	-0.137	1.364	0.189
		(0.144)		(0.226)		(0.178)
Total score Rosenberg Scale	32.991	0.101	32.777	0.119	33.306	0.081
		(0.154)		(0.224)		(0.212)
Total score Grit Scale	43.707	-0.248	43.340	-0.171	44.251	-0.325
		(0.198)		(0.285)		(0.274)
Multivariate F p-value		0.756		0.479		0.594

Notes: This table reports balance checks of being assigned to more central peers on baseline characteristics. All regressions include strata fixed effects and include the higher-achieving peers treatment. For the 2017 cohort, all regressions include gender-by-classroom fixed effects. The control group is defined as being assigned to less central and lower-achieving peers. The "F p-value" corresponds to the F-statistic of the more central peers treatment of multivariate regressions that include all the variables at baseline. Standard errors are clustered at the peer-group-type-by-student-type level; *** p-value<0.01, ** p-value<0.05, * p-value<0.1.

Table A.3: Balance Tests for the Higher-Achieving Peers Treatment

Variable	All Students		Lower-achieving at Baseline		Higher-achieving at Baseline	
	Control mean	Difference	Control mean	Difference	Control mean	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
Admission test	-0.163	0.022	-0.787	0.004	0.764	0.036
		(0.015)		(0.018)		(0.026)
Interview score	0.049	-0.016	0.212	-0.031	-0.192	-0.015
		(0.024)		(0.031)		(0.041)
Social-fit score	0.045	-0.030	0.164	-0.014	-0.130	-0.050
		(0.022)		(0.029)		(0.039)
Female (%)	0.567	0.000	0.566	-0.000	0.567	0.000
		(0.000)		(0.000)		(0.000)
Poor (%)	0.418	0.010	0.456	0.015	0.362	-0.002
		(0.012)		(0.018)		(0.017)
Rural household (%)	0.267	0.004	0.308	-0.019	0.206	0.017
		(0.011)		(0.018)		(0.015)
Subsidized health insurance	0.504	0.013	0.517	0.045	0.485	-0.019
		(0.015)		(0.021)		(0.022)
Math scores	-0.081	0.007	-0.345	-0.023	0.310	0.028
		(0.019)		(0.026)		(0.030)
Reading scores	-0.047	-0.014	-0.234	-0.033	0.229	0.006
		(0.018)		(0.027)		(0.027)
Social skills	-0.015	-0.017	-0.090	-0.029	0.097	-0.006
		(0.022)		(0.030)		(0.032)
Degree friends	8.023	-0.438	7.787	-0.418	8.374	-0.457
		(0.134)		(0.183)		(0.195)
Centrality friends	0.029	-0.081	-0.025	-0.082	0.107	-0.080
		(0.025)		(0.033)		(0.037)
Degree study	4.719	0.012	4.609	-0.145	4.882	0.167
		(0.069)		(0.097)		(0.096)
Centrality study	-0.011	-0.004	-0.058	-0.028	0.059	0.020
		(0.027)		(0.034)		(0.042)
Degree all	11.125	-0.244	10.946	-0.394	11.392	-0.094
		(0.137)		(0.187)		(0.201)
Centrality all	-0.002	-0.012	-0.031	-0.036	0.041	0.011
		(0.018)		(0.025)		(0.025)
Reading the mind in the eyes	20.602	-0.264	20.248	-0.231	21.127	-0.295
		(0.128)		(0.180)		(0.183)
Peers' perception leadership	2.487	0.011	1.923	-0.026	3.328	0.047
		(0.148)		(0.178)		(0.236)
Peers' perception friendliness	2.688	0.011	2.612	-0.002	2.801	0.024
		(0.078)		(0.112)		(0.109)
Peers' perception popularity	2.361	-0.018	2.006	-0.025	2.890	-0.014
		(0.158)		(0.185)		(0.255)
Peers' perception shyness	2.017	0.038	2.108	0.162	1.881	-0.086
		(0.146)		(0.211)		(0.203)
Total score Rosenberg Scale	33.013	0.138	32.854	0.127	33.249	0.153
		(0.154)		(0.221)		(0.215)
Total score Grit Scale	43.568	0.160	43.330	0.431	43.921	-0.108
		(0.196)		(0.271)		(0.282)
Multivariate F p-value		0.256		0.889		0.232

Notes: This table reports balance checks of being assigned to higher-achieving peers on baseline characteristics. All regressions include strata fixed effects, control for the baseline value of the dependent variable, and include the more central peers treatment. For the 2017 cohort, all regressions include gender-by-classroom fixed effects. The control group is defined as being assigned to less central and lower-achieving peers. The "F p-value" corresponds to the F-statistic of the higher-achieving peers treatment of multivariate regressions that include all the variables at baseline. Standard errors are clustered at the peer-group-type-by-student-type level level; *** p-value<0.01, ** p-value<0.05, * p-value<0.1.

Table A.4: Correlations between Types of Skills and Longer-term Outcomes

Dependent variable:	Dropout	College	Certified	Top 20
	(1)	(2)	(3)	(4)
Social skills	-0.008***	0.024***	0.036***	0.026***
	(0.001)	(0.008)	(0.008)	(0.008)
Math scores	-0.002	0.069***	0.079***	0.090***
	(0.001)	(0.010)	(0.010)	(0.010)
Reading scores	0.002*	0.019**	0.025***	0.036***
	(0.001)	(0.009)	(0.009)	(0.009)
mean control	0.02	0.62	0.33	0.27
N	6,147	3,654	3,654	3,654

Notes: This table reports the correlation of social skills with longer-term outcomes. The three variables of interest are standardized. All models include school-by-cohort-by-gender fixed effects. Column 1 presents the results on the dropout rate with data available for all cohorts. Columns 3 to 4 present the estimates on college outcomes only available for the 2015-16 cohorts. Robust standard errors in parentheses; *** p-value<0.01, ** p-value<0.05, * p-value<0.1.

Table A.5: Randomization Inference

Group	Treatment	Dependent variables			
(1)	(2)	(3)	(4)	(5)	(6)
		Panel A: Results on Social Skills			
		Connections	Centrality	Psychological Tests	Peers' perception
All students	More central peers	0.003	0.012	0.070	0.030
		[0.978]	[0.623]	[0.006]	[0.142]
	Higher-achieving peers	-0.117	-0.001	-0.017	0.012
		[0.378]	[0.960]	[0.409]	[0.479]
	Joint test	[0.668]	[0.878]	[0.023]	[0.232]
Boys	More central peers	0.498	0.099	0.144	0.057
		[0.038]	[0.011]	[0.002]	[0.092]
	Higher-achieving peers	0.042	0.032	-0.028	-0.017
		[0.834]	[0.262]	[0.399]	[0.546]
	Joint test	[0.115]	[0.027]	[0.003]	[0.222]
Less central boys	More central peers	0.946	0.195	0.237	0.099
		[0.004]	[0.001]	[0.001]	[0.010]
	Higher-achieving peers	-0.232	0.009	0.052	-0.090
		[0.466]	[0.865]	[0.383]	[0.021]
	Joint test	[0.012]	[0.001]	[0.001]	[0.005]
		Panel B: Results on Academic Skills			
		Grades		Test Scores	
		Math	Reading	Math	Reading
All students	More central peers	0.022	0.041	-0.022	0.025
		[0.517]	[0.246]	[0.297]	[0.365]
	Higher-achieving peers	0.009	-0.018	-0.027	-0.033
		[0.691]	[0.442]	[0.125]	[0.102]
	Joint test	[0.742]	[0.398]	[0.184]	[0.201]
Girls	More central peers	0.000	0.044	-0.020	-0.007
		[0.997]	[0.416]	[0.449]	[0.865]
	Higher-achieving peers	-0.012	-0.006	-0.027	-0.073
		[0.677]	[0.834]	[0.165]	[0.005]
	Joint test	[0.906]	[0.686]	[0.317]	[0.026]
Lower-achieving	More central peers	0.013	0.067	-0.015	0.041
		[0.790]	[0.183]	[0.631]	[0.295]
	Higher-achieving peers	-0.061	-0.075	-0.043	-0.041
		[0.075]	[0.039]	[0.098]	[0.170]
	Joint test	[0.220]	[0.060]	[0.223]	[0.266]
Lower-achieving girls	More central peers	-0.010	0.063	-0.061	0.020
		[0.858]	[0.369]	[0.081]	[0.709]
	Higher-achieving peers	-0.115	-0.068	-0.069	-0.080
		[0.010]	[0.137]	[0.015]	[0.048]
	Joint test	[0.044]	[0.277]	[0.016]	[0.161]

Notes: This table reports randomization inference p-values for social and academic outcomes by groups of students and treatments. All the estimates come from a separate regression for each subgroup. The first column presents the group for which the test is performed, and the second column the respective treatment. Columns 3 to 6 show the set of outcomes: social skills outcomes in Panel A and academic outcomes in Panel B. All p-values are in square brackets. The "Joint test" corresponds to the p-value of the joint test of at least one treatment being statistically significant. The p-values were calculated using the procedure developed by Young (2018) with 1,000 randomization iterations.

Table A.6: Multiple-Hypotheses Testing

Group	Test	Dependent variables			
(1)	(2)	(3)	(4)	(5)	(6)
		Panel A: Results on Social Skills			
		Connections	Centrality	Psychological Tests	Peers' perception
Boys	Point estimate	0.510	0.100	0.144	0.061
	Sidak and Holm	0.060	0.031	0.002	0.118
	Bonferroni and Holm	0.061	0.031	0.002	0.122
	Westfall and Young	0.107	0.058	0.004	0.188
Less sociable boys	Point estimate	0.946	0.200	0.237	0.099
	Sidak and Holm	0.012	0.002	0.001	0.032
	Westfall and Young	0.012	0.002	0.001	0.032
		0.029	0.004	0.002	0.079
		Panel B: Results on Academic Skills			
		Grades		Test Scores	
		Math	Reading	Math	Reading
Lower-achieving	Point estimate	-0.061	-0.075	-0.043	-0.041
	Sidak and Holm	0.176	0.083	0.172	0.348
	Bonferroni and Holm	0.184	0.085	0.180	0.385
	Westfall and Young	0.259	0.135	0.284	0.414
Girls	Point estimate	-0.012	-0.006	-0.027	-0.073
	Sidak and Holm	0.686	0.842	0.316	0.011
	Bonferroni and Holm	0.686	0.842	0.346	0.012
	Westfall and Young	0.749	0.880	0.400	0.018
Lower-achieving girls	Point estimate	-0.115	-0.068	-0.069	-0.080
	Sidak and Holm	0.063	0.422	0.055	0.193
	Bonferroni and Holm	0.065	0.512	0.056	0.209
	Westfall and Young	0.113	0.567	0.103	0.262

Notes: This table reports multiple-hypotheses testing p-values. The first column presents the group for which the test is performed among all the possible classifications of students. For instance, if the groups are boys, the reported test is on the treatment effect for boys of multiple hypotheses that considers the impact on both boys and girls. The second column corresponds to the respective test of multiple hypotheses. Columns 3 to 6 show the set of outcomes: social skills outcomes in Panel A and academic outcomes in Panel B. Calculations were performed using the *wyoung* command developed by Jones et al. (2019).

Table A.7: Treatment Effects on Social Skills (with an Interaction Term)

Group:	All students			Less central			More central		
	All	Boys	Girls	All	Boys	Girls	All	Boys	Girls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
More central	0.052	0.121**	0.002	0.065	0.236***	-0.042	0.060	0.023	0.081
	(0.034)	(0.056)	(0.044)	(0.046)	(0.070)	(0.060)	(0.053)	(0.091)	(0.065)
Higher-achieving	0.039	0.021	0.048	-0.046	-0.006	-0.079	0.161***	0.045	0.232***
	(0.034)	(0.056)	(0.043)	(0.042)	(0.067)	(0.052)	(0.058)	(0.101)	(0.069)
Interaction	-0.059	-0.009	-0.095	-0.006	-0.048	0.022	-0.146**	0.015	-0.253***
	(0.049)	(0.077)	(0.064)	(0.067)	(0.105)	(0.084)	(0.074)	(0.121)	(0.094)
N	3,654	1,490	2,164	1,832	753	1,079	1,822	737	1,085

Notes: This table reports the effect of being assigned to more central peers, higher-achieving peers, and the interaction of both treatments on social skills. All regressions control for strata fixed effects and selected covariates at baseline, including the score on the admission test, math and reading scores, centrality and degree in the social network, peers' perception of social skills, and own perception of academic and social skills. The sample includes students from the 2015-16 cohorts as there is no information on centrality at baseline for the 2017 cohort. Standard errors are clustered at the peer-group-type-by-student-type level; *** p-value<0.01, ** p-value<0.05, * p-value<0.1.

Table A.8: Treatment Effects on Academic Achievement (with an Interaction Term)

Group:	All students			Lower-achieving			Higher-achieving		
	All	Boys	Girls	All	Boys	Girls	All	Boys	Girls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Dependent variable math scores									
More central	-0.028	-0.061	-0.004	-0.011	0.033	-0.047	-0.061	-0.190**	0.030
	(0.027)	(0.047)	(0.033)	(0.037)	(0.066)	(0.041)	(0.044)	(0.073)	(0.054)
Higher-achieving	-0.031	-0.045	-0.017	-0.039	-0.017	-0.057*	-0.046	-0.096*	-0.002
	(0.020)	(0.034)	(0.023)	(0.031)	(0.057)	(0.032)	(0.031)	(0.052)	(0.038)
Interaction	0.012	0.073	-0.031	-0.011	0.032	-0.033	0.056	0.160*	-0.018
	(0.036)	(0.061)	(0.044)	(0.056)	(0.095)	(0.065)	(0.052)	(0.089)	(0.064)
N	5,681	2,505	3,176	2,778	1,236	1,542	2,890	1,259	1,631
Panel B: Dependent variable reading scores									
More central	0.041	0.113**	-0.014	0.015	0.089	-0.030	0.074	0.161*	0.004
	(0.035)	(0.052)	(0.048)	(0.048)	(0.070)	(0.066)	(0.056)	(0.084)	(0.075)
Higher-achieving	-0.023	0.047	-0.078**	-0.065	0.014	-0.125**	0.005	0.080	-0.055
	(0.024)	(0.037)	(0.031)	(0.040)	(0.061)	(0.053)	(0.036)	(0.056)	(0.045)
Interaction	-0.031	-0.093	0.015	0.063	-0.032	0.119	-0.110*	-0.164	-0.059
	(0.043)	(0.066)	(0.058)	(0.065)	(0.101)	(0.087)	(0.066)	(0.101)	(0.087)
N	5,796	2,540	3,256	2,860	1,260	1,600	2,923	1,270	1,653

Notes: This table reports the effect of being assigned to more central peers, higher-achieving peers, and the interaction of both treatments on academic achievement. All regressions control for strata fixed effects and selected covariates at baseline, including the score on the admission test, math and reading scores, centrality and degree in the social network, peers' perception of social skills, and own perception of academic and social skills. For the 2017 cohort, all regressions include gender-by-classroom fixed effects. Standard errors are clustered at the peer-group-type-by-student-type level; *** p-value<0.01, ** p-value<0.05, * p-value<0.1.

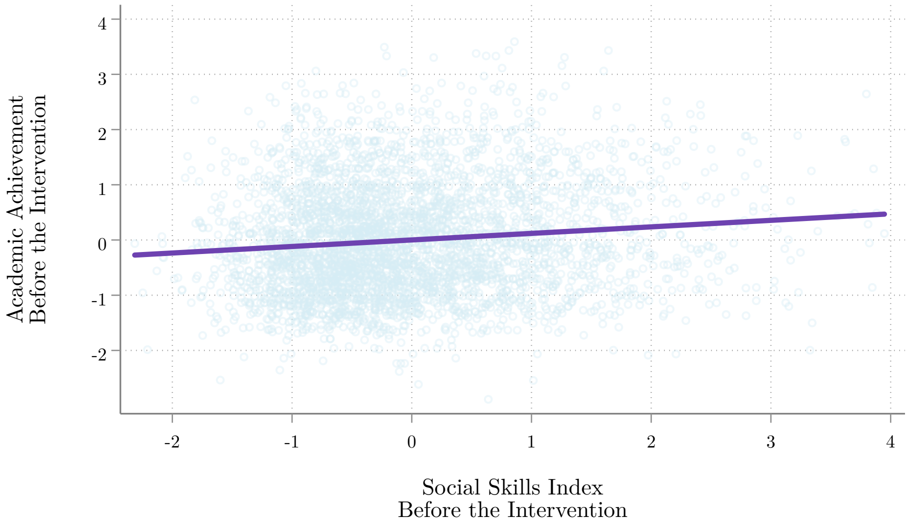
Table A.9: Social Connections with Neighbors

Dependent variable: (network)	Friend	Study	Social	Any	Help academic	Help personal
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Less central students at baseline						
More central	0.002	-0.010	-0.032	0.018	-0.000	-0.036
	(0.046)	(0.033)	(0.042)	(0.053)	(0.032)	(0.034)
Higher-achieving	-0.051	-0.021	-0.064	-0.027	0.059*	0.027
	(0.043)	(0.033)	(0.041)	(0.051)	(0.033)	(0.035)
More central × boy	0.035	0.007	0.076	0.051	0.033	0.075*
	(0.065)	(0.050)	(0.062)	(0.072)	(0.046)	(0.044)
Higher-achieving × boy	0.054	0.014	0.067	0.021	-0.017	-0.017
	(0.065)	(0.050)	(0.060)	(0.072)	(0.045)	(0.044)
mean control	0.57	0.39	0.54	0.74	0.20	0.28
p-val mc boys	0.426	0.936	0.328	0.159	0.317	0.156
p-val ha boys	0.954	0.850	0.947	0.909	0.180	0.721
N	1,829	1,829	1,829	1,829	1,829	1,829
Panel B: Lower-achieving students at baseline						
More central	-0.126*	-0.052	-0.105	-0.061	0.009	-0.022
	(0.065)	(0.050)	(0.068)	(0.079)	(0.048)	(0.054)
Higher-achieving	0.041	0.055	-0.014	0.034	0.092**	0.106**
	(0.053)	(0.046)	(0.054)	(0.058)	(0.038)	(0.046)
More central × boy	0.199**	0.037	0.141	0.122	0.008	0.041
	(0.090)	(0.073)	(0.089)	(0.104)	(0.068)	(0.064)
Higher-achieving × boy	0.014	-0.051	-0.116	-0.022	-0.027	-0.151***
	(0.082)	(0.067)	(0.076)	(0.085)	(0.055)	(0.055)
mean control	1.15	0.84	1.04	1.45	0.47	0.51
p-val mc boys	0.255	0.781	0.560	0.377	0.743	0.613
p-val ha boys	0.381	0.946	0.016	0.857	0.103	0.147
N	2,269	2,269	2,269	2,269	2,269	2,269

Notes: This table reports the effect of being assigned to more central and higher-achieving peers on the number of social connections with neighbors in dormitories. All regressions control for strata fixed effects and selected covariates at baseline, including the score on the admission test, math and reading scores, centrality and degree in the social network, peers' perception of social skills, and own perception of academic and social skills. For the 2017 cohort, all regressions include strata-by-classroom fixed effects. The control group is defined as being assigned to less central and lower-achieving peers. Standard errors are clustered at the peer-group-type-by-student-type level; *** p-value<0.01, ** p-value<0.05, * p-value<0.1.

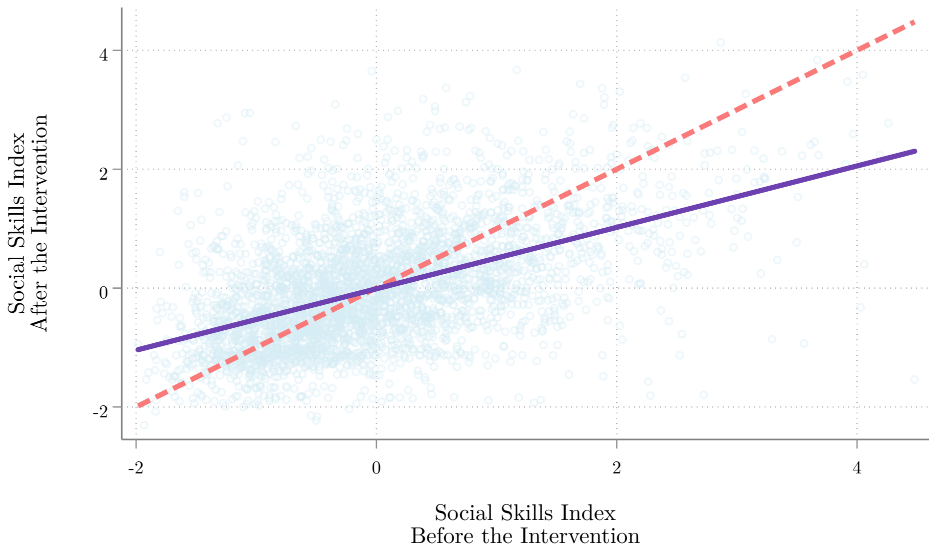
Figure A.1: Social Skills Index

Panel A: Correlation with Academic Achievement



achievement=0.00+0.11*socialskills

Panel B: Correlation over Time

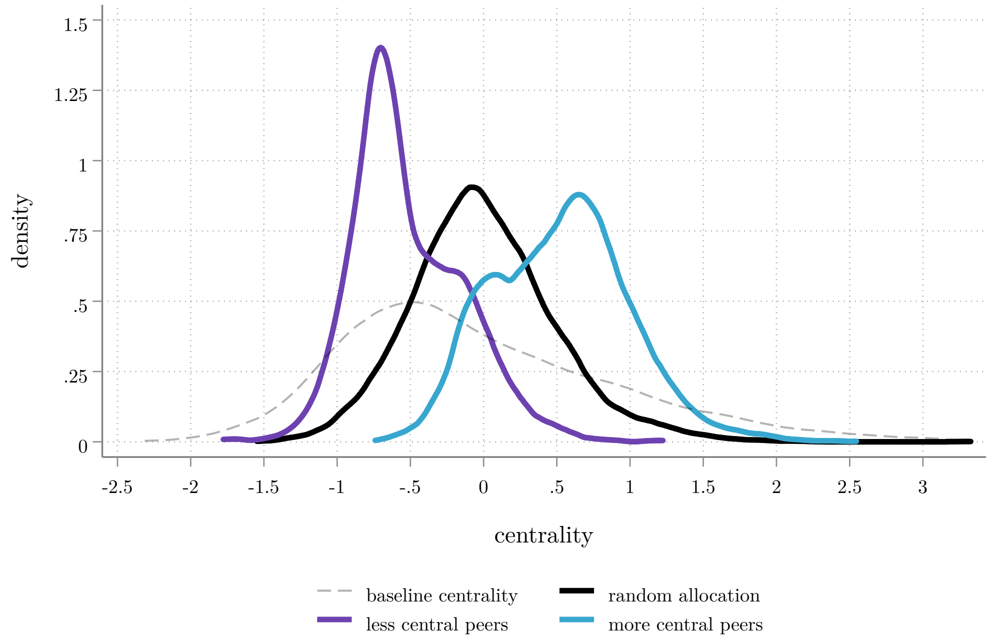


$\text{socialskills}_t = -0.01 + 0.52 * \text{socialskills}_{t-1}$

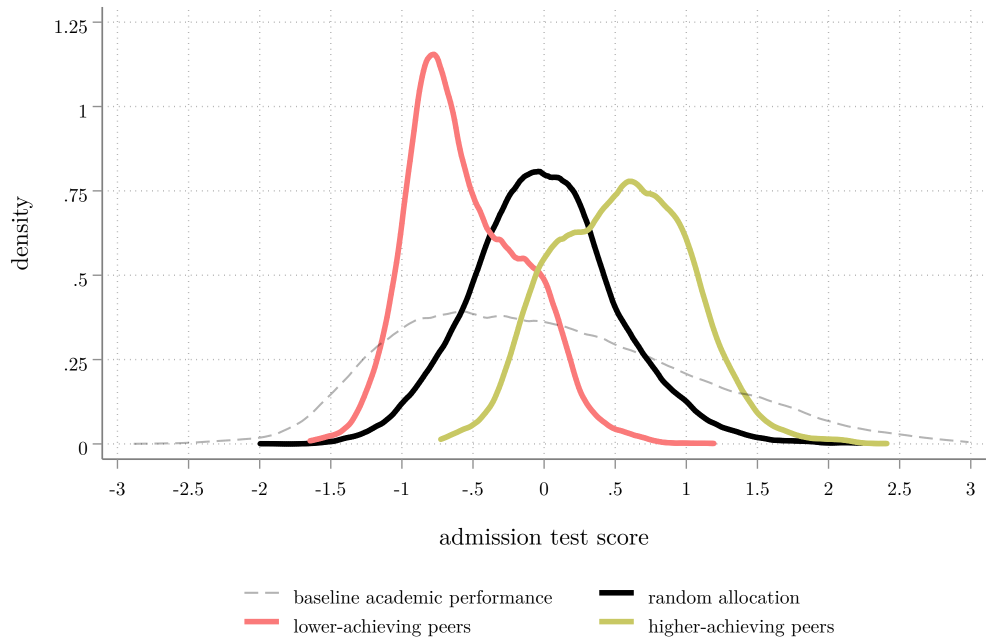
Notes: Panel A shows a scatter plot of academic achievement and the social skills index at baseline for the 2015-16 cohorts by student type. A one-standard-deviation of the social skills index predicts an increase in 0.11 standard deviations of academic achievement at baseline. Panel B shows a scatter plot and the linear prediction of the social skills index before and after the intervention. A one-standard-deviation of the social skills index before the intervention predicts an increase of 0.42 in the social skills index after the intervention.

Figure A.2: Distribution of Baseline and Peer Characteristics

Panel A: Social Centrality

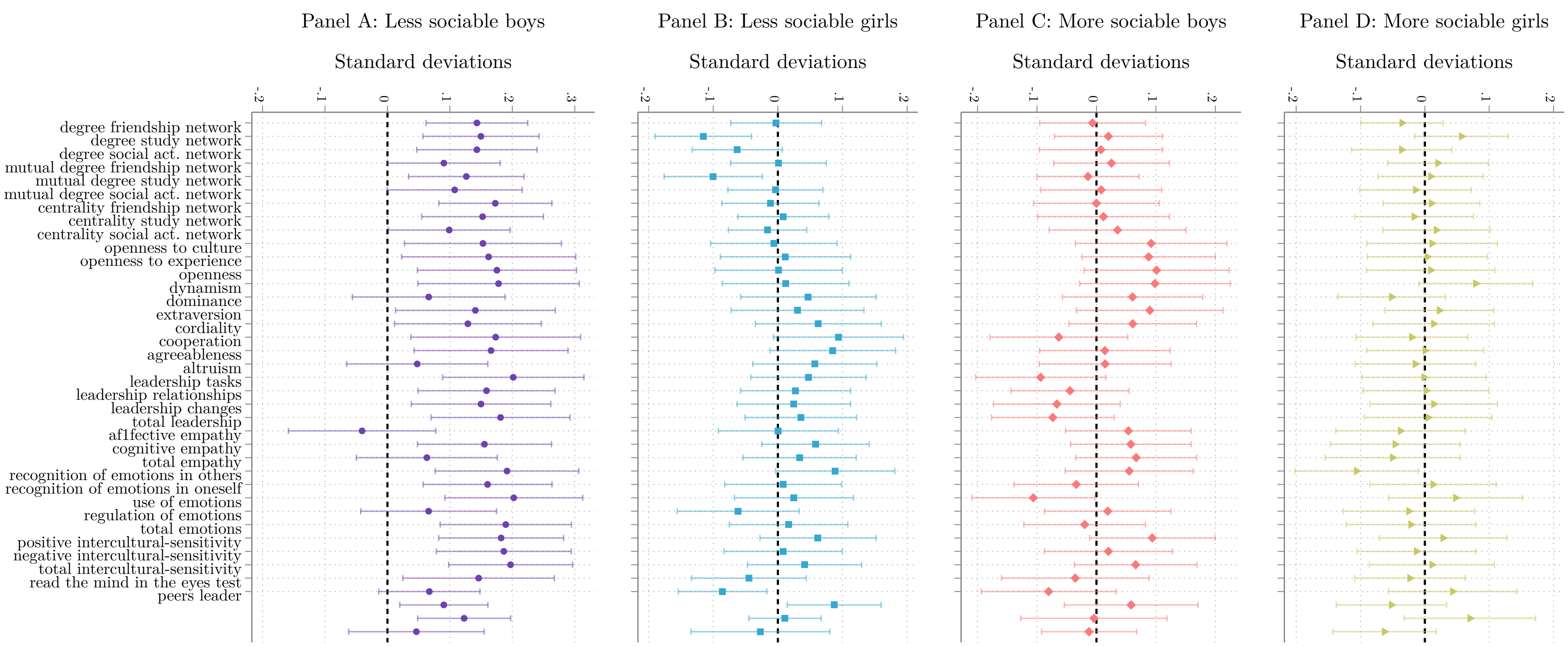


Panel B: Academic Achievement



Notes: This figure plots the distribution of baseline and peer characteristics in the allocation to the student-peer type combinations. It also shows the distribution of peer characteristics using random assignment to groups for comparison.

Figure A.3: Effects of More Central Peers on Social Skills



Notes: This figure reports treatment effects and 90% confidence intervals of being assigned to more central peers on social skills outcomes. All regressions include strata fixed effects and control for the baseline value of the dependent variable. The control group is defined as being assigned to less central peers. Standard errors are clustered at the peer-group-type-by-student-type level.

B Estimation of Peer Effects

This appendix describes the methodological concerns about exploiting random allocation to groups to identify peer effects and how my experimental design addresses them.

B.1 Random Allocation to Groups

A widely used research method to estimate peer effects - is to exploit random allocation to groups. In settings where schools and colleges apply this method, there is no self-selection of peers, and ex-ante individual and peer characteristics are unrelated. Hence, random group allocation allows researchers to estimate the causal impact of predetermined peer characteristics on individual outcomes.

However, there are some problems with using random allocation to groups to estimate peer effects, which can either be under- or over-estimated. By construction, the variation in peer characteristics from random groups is small. As groups get larger, this problem is aggravated. As Manski in Epple and Romano (2011) comments, "Random assignment will not work well in a large group setting, because all groups will have essentially the same distribution of types." Similarly, Angrist (2014) argues that "the interpretation of results from models that rely solely on chance variation in peer groups is therefore complicated by bias from weak instruments."

I now build on Angrist (2014) to describe methodological concerns about random allocation to groups. I then relate this to the existing literature and explain the advantages of my experimental design.

To introduce the problem, consider the following peer-effects model:

$$y_{ig} = \alpha + \pi_0 x_{ig} + \pi_1 \overline{x}_g + \varepsilon_{ig} \tag{B.1}$$

where y_{ig} is the outcome of individual i when assigned to group g, x_{ig} is a pre-specified exogenous characteristic of individual i in group g, and $\overline{x}_g$ is the mean of the exogenous characteristic x among those in group g.

Parameter π_1 is the causal effect of a change in the group average of x over students' outcomes. Acemoglu and Angrist (2000) show that equation B.1 relates to whether a 2SLS estimator using group dummies to instrument individual characteristics differs from OLS estimates of the effect of these characteristics. Specifically:

$$\pi_1 = \frac{\psi_1 - \psi_0}{1 - \tau^2}, \tag{B.2}$$

where ψ_0 is the OLS estimator of the parameter ψ in the following model:

$$y_{ig} = \alpha + \psi x_{ig} + \varepsilon_i, \tag{B.3}$$

and ψ_1 is the 2SLS estimator of this model, using the vector of group dummies g as instruments for x_i. The parameter τ^2 is the first-stage R-squared associated with this 2SLS estimate; the variation in x_i explained by the group dummies.

Angrist (2014) argues that because of the relationship in equation B.2, the estimation of peer effects using random allocation to groups can suffer from weak instruments. Furthermore, even with systematic variation in group composition, the 2SLS estimates can exceed the OLS estimates for other reasons unrelated to social effects, such as measurement error. The use of variation across groups to estimate peer effects can confound peer effects with factors unrelated to social influences.

Nevertheless, as pointed out by Feld and Zölitz (2017), Angrist (2014) does not explicitly show under what conditions an upward bias exists and how it depends on the underlying parameters of the model. In fact, under regular conditions, 2SLS estimates with weak instruments are biased towards OLS, which implies that π_1 tends towards zero. Feld and Zölitz (2017) also show that with classical measurement error in the exogenous characteristics x, and with a random group assignment, peer-effects estimates are biased towards zero. In this vein, exploiting random allocation to groups seems to underestimate rather than overestimate peer effects.

Still, the evidence in the literature suggests that estimates of peer effects increase with group size when the variation in peer characteristics weakens. For instance, with an average classroom size of forty-four students (ranging from nineteen to ninety-one), Duflo et al. (2011) find that a one–standard deviation increase in average peer test scores would increase the test score of a student by 0.445 standard deviations, an effect they claim is comparable to previous findings. Similarly, in Carrell et al. (2009), a one-hundred-point increase in peer SAT verbal scores has negligible peer effects on grades when roommates are the relevant peer group (0.003 (s.e. 0.019)) but sizable and significant effects when the peers are other freshmen in the squadron (0.338 (s.e. 0.107)), where group size is larger.[22] Carrell et al. (2013) use the last set of estimates in a posterior experiment that estimates the effect of optimal groups. Contrary to the prediction, they find a negative treatment effect. While the authors attribute this disappointing result to the endogenous patterns of social interactions, Angrist (2014) argues that it might be driven in part by the imprecision of a 2SLS design without a real first stage.

Epple and Romano (2011) reach a similar conclusion with respect to group size. Their handbook chapter concludes that a one-unit increase in peer average ability increases a student's achievement by 0.2 to 0.6 points. Epple and Romano (2011) also consider it surprising that studies that exploit randomization tend to find larger peer effects than those typically found with other identification strategies. For instance, studies using quasi-experimental variation such as Dobbie and Fryer (2014) and Abdulkadiroğlu et al. (2014) find little evidence of peer effects on test scores and college outcomes. Likewise, in the same context of the large peer effects mentioned above, Duflo et al. (2011) find little evidence of peer effects when exploiting an RD on the median student of a tracking system.

This pattern is not limited to classrooms or groups of very large size. Garlick (2018) estimates an impact of 0.216 s.d. when students are assigned to dorms with an average size of 128 students in a South African university. Glaeser et al. (2003) find that the impact on fraternity participation of the average fraction of peers that drink in high school increases with the size of the reference group, even when the differences are small (see Table 1 in Angrist (2014)). While the estimated effect is 0.098 at the dorm level (average size of 2.3 students), it increases by 50% to 0.145 at the floor level (average size of 8.0 students). The impact is even larger at the building level (0.232), where the average group size is 28 students.

Two explanations for this phenomenon can be extracted from equation B.2. The first one is that as groups get larger and the variation in peer characteristics gets weaker, estimates of peer effects become more imprecise. To see this, notice that the variance of the estimator of π_1 in

[22] A squadron comprises approximately 120 students (freshmen through seniors).

equation 1 is given by the following:

$$var(\hat{\pi}_1) = \frac{1}{N_s}\,\frac{N_g}{N_g - 1}\,\frac{\sigma^2_{\varepsilon}}{var(\overline{x}_g)} \tag{B.4}$$

where N_s is the sample size, $var(\overline{x}_g) = \frac{\sigma^2_x}{N_g}$, and N_g is the group size. The variance of π_1 (equation B.4) is an increasing function in $N_g \geq 2$. Intuitively, as groups get larger, the variation in peer characteristics is lower and hence the precision of the estimate decreases. This argument has been previously explored by Angrist (2014) as an explanation for the estimates in Glaeser et al. (2003) and the differences between Carrell et al. (2009) and Carrell et al. (2013).

The second and less explored explanation of positive peer effects that increase with group size is the amplification of bias when the variation in peer characteristics is weak. This is a similar situation to the one encountered when instruments explain little of the variation in the endogenous variables. A very small violation of the exclusion restriction can lead to a large (asymptotic) bias. Following equation B.2, this would imply that estimates of peer effects grow with group size, as the difference between 2SLS and OLS estimates is increasing.

The probability limit of ψ_1 is:

$$plim\ \psi_1 = \pi_0 + \frac{cov(\varepsilon_{ig}, \overline{x}_g)}{var(\overline{x}_g)}.$$

As groups get larger and the variance of $\overline{x}_g$ decreases, any correlation between the error term and the average peer characteristics is amplified. Notice that this is the case even if the covariance between the error term and $\overline{x}_g$ also decreases with group size but at a lower rate than the decrease in the variance of $\overline{x}_g$. Any model with this feature will amplify the bias with group size much like weak instruments do.

Figure B.1 introduces simulations of the linear-in-means peer effects model (equation 1), illustrating both problems. In particular, the left column in Panel A presents the distribution of the estimates of π_1 in equation 1, assuming that $\pi_1 = 0$. In general, and as expected from equation B.4, estimates become imprecise as the group size increases. However, these losses in precision imply that we should observe both large positive and large negative estimates across studies, which is inconsistent with the empirical evidence.

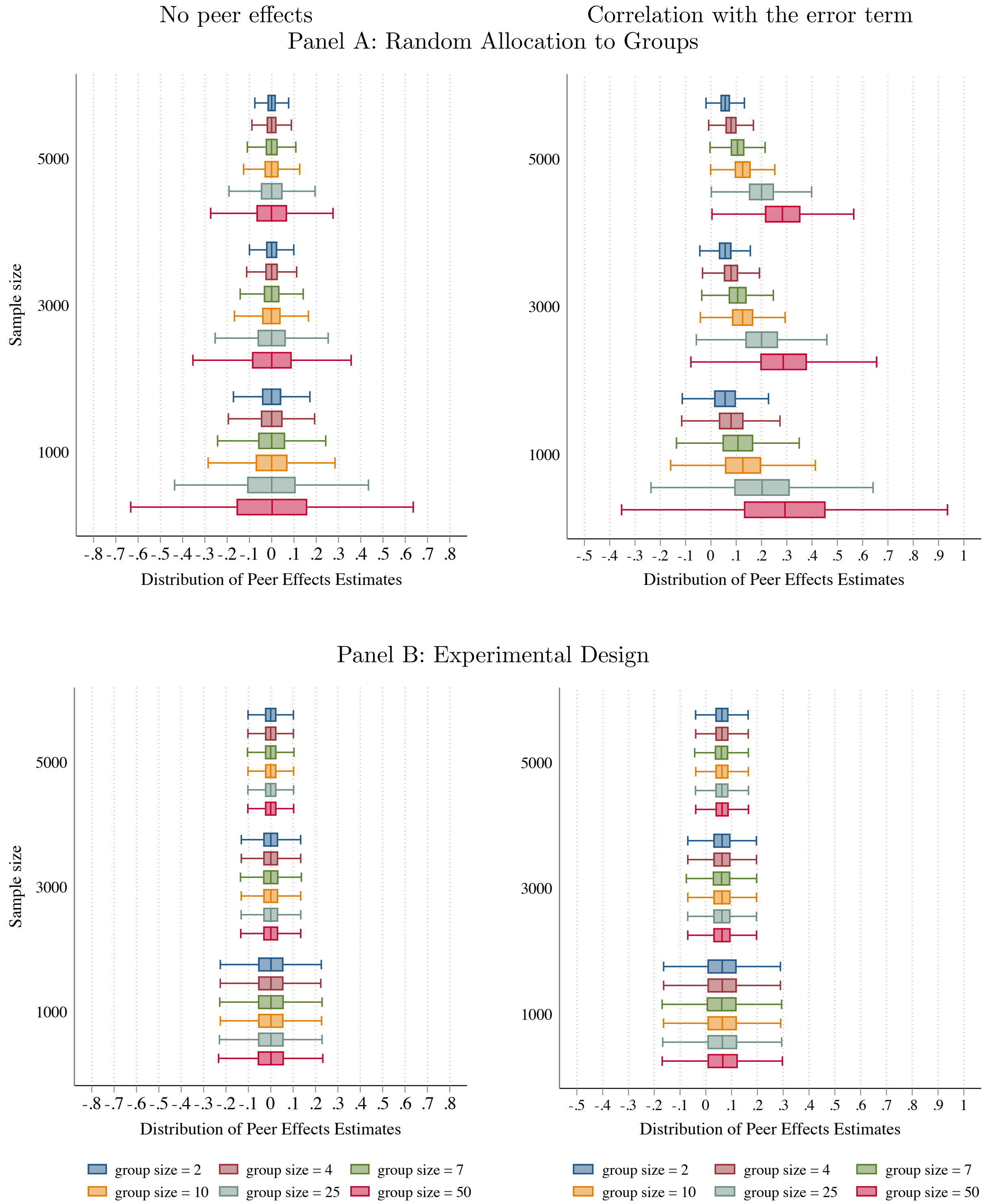
A second explanation for the increase in peer effects estimates when group size is larger is the amplification of bias. The right plot in Panel A of Figure B.1 illustrates this concern. For this plot, I consider a small nonlinear peer effect in the error term. In particular, all groups with an average score above the median receive a positive shock of 0.1 –a relatively small nonlinearity. As the plot shows, the misspecification of the functional form amplifies the bias of the linear-in-means estimate when the group size increases.[23] While the average bias is only about 0.056 when the groups are pairs, it rapidly grows to 0.082 when the group size is four. The increase in the magnitude of the bias is explosive. With a group size of seven students, the bias is twice as large as the one with two students. A larger group size of twenty-five students quadruples the bias, with an average and median estimate of 0.20.

My experimental design does not lose precision or increase bias with group size. This is because there is substantial variation in peer characteristics by virtue of the treatment arms and because

[23] In an individual model, this correlation with the error term would generate a bias of $\hat{\pi}_1$ of 0.05. However, as illustrated by the plot, the bias amplifies with the group size.

the identification strategy does not rely on variation across groups. Parameter π_1 is estimated via a 2SLS model with a single instrument and a strong first stage. Simulations in Panel B of Figure B.1 numerically illustrate how my experimental design preserves precision and limits the bias of estimates with different group sizes. In the left panel, I assume that $\pi_1 = 0$ and plot the corresponding peer-effects estimates. The precision of the estimates remains constant with group size and only varies with the sample size. Similarly, in the right panel, I consider a nonlinear correlation with the error term, but the positive bias remains constant regardless of group size. Both plots sharply contrast with the findings from a random allocation to groups.

Figure B.1: Simulations of Peer Effects



Notes: Monte Carlo simulations based on 10,000 repetitions of the estimate of parameter π_1 in equation B.3. The simulations assume that $x_i \sim N(0,1)$, and $\nu_{ig} \sim N(0,1)$. In the left column, $\pi_0 = 1$, $\pi_1 = 0$, and $\varepsilon_{ig} = \nu_{ig}$. In the right column, $\pi_0 = 1$, $\pi_1 = 0$, and $\varepsilon_{ig} = 0.1 \times I(\overline{x}_g \geq 0) + \nu_{ig}$.

C Experimental Design: An Application

To understand how the design works in practice, consider an example similar to the one described by Guryan et al. (2009). I use this example to explain why my design guarantees strong variation and is not subject to exclusion bias.

In this example, twelve individuals are randomly assigned to groups of four. To easily navigate through this example, let the individuals have pre-determined levels of an attribute x_i, where $x_i = i$ and i has a range from 1 (the lowest skill level) to 12 (the highest skill level). Following the first step of my research design, individuals 1-6 are classified as low type and individuals 7-12 as high type. Second, these individuals are randomly assigned to the high-type peer treatment:

- Four of the individuals 7 to 12 are randomly assigned to high-type peers and put together in the same group. This is Combination A composed of high-type students assigned to high-type peers.
- The other two high-type individuals are allocated to two low-type individuals (selected individuals in the 1-6 range) who were randomly assigned to the high-peers treatment. This is Combination B, a mixed combination of high- and low-type peers.
- The remaining four low-type individuals from 1 to 6 were assigned to the low-type peer treatment and belong to the same group. This is Combination C, composed of low-type students assigned to low-type peers.

From our simple example above, we have four students in A, four in B, and four in C. As groups are of size four, all students will have three peers in their group.

Let's move now to step 3 of the research design, focusing on the strength of the first stage and the exclusion bias. I will do this for low-type students. There will only be low types in Combinations B and C. For those assigned to the treatment (Combination B), two out of three peers are randomly chosen from individuals 7-12. For a low-type in Combination B, only one out of 3 peers will come from individuals 1 to 6 (low-types too). The expected value of the leave-out mean for low-types who are assigned to the treatment (Combination B) can then be described by:

$$\mathbb{E}\left[\overline{x}_{g,-i}|i, H_i = 1\right] = \underbrace{\frac{2}{3}\left(\frac{\sum_{j=7}^{12} j}{6}\right)}_{\substack{\text{peer variation coming}\\ \text{from being treated}}} + \underbrace{\frac{1}{3}\left(\frac{\sum_{j=1}^{6} j}{5} - \frac{i}{5}\right)}_{\substack{\text{same-type peer}\\ \text{assigned to the same group}}}, \tag{C.1}$$

In Combination C, low-type students are assigned to the control group; all three peers come from individuals 1 to 6. In this case, the expected value of the leave-out mean for low-types who are assigned to the control (Combination C) can be described by:

$$\mathbb{E}\left[\overline{x}_{g,-i}|i, H_i = 0\right] = \underbrace{\frac{2}{3}\left(\frac{\sum_{j=1}^{6} j}{5} - \frac{i}{5}\right)}_{\substack{\text{peer variation coming}\\ \text{from being in the control group}}} + \underbrace{\frac{1}{3}\left(\frac{\sum_{j=1}^{6} j}{5} - \frac{i}{5}\right)}_{\substack{\text{same-type peer}\\ \text{assigned to the same group}}} \tag{C.2}$$

Then, for individual i, the expected value of the difference in peer characteristics, conditional on being treated (equation C.1) and in the control group C.2 is:

$$\lambda_i = \mathbb{E}\left[\overline{x}_{g,-i}|i, H_i = 1\right] - \mathbb{E}\left[\overline{x}_{g-i}|i, H_i = 0\right] \quad = \tfrac{53-2i}{15} \tag{C.3}$$

Note how the second term from expressions C.1 and C.2 cancel out together. Hence, any expected difference between treatment and control groups will arise from the first term (the peers that provide the treatment or the control). We can now apply the law of iterated expectations to calculate the strength of the first stage for low-type individuals (the parameter λ in equation 2:

$$\lambda = \mathbb{E}_i\left[\lambda_i\right] = \frac{53 + 2\mathbb{E}\left[i\right]}{15} = \frac{60}{15} = 4, \tag{C.4}$$

where the expected value of the attribute is $\mathbb{E}\left[i\right] = \frac{7}{2}$, as attributes from low-type individuals range from 1 to 6.

There are two things to notice from equation C.4. First, not only is the difference of 4 driven by the treatment assignment, as previously mentioned, but it is different from the near-zero result we would have found if the variation in peer characteristics were weak. A first stage of 4 is also strong and larger than a one-standard deviation in the distribution of attribute x_i. Second, while expressions C.1 and C.2 are subject to the exclusion bias (term $\frac{i}{5}$), the main source of identification comes from variation across treatment and control groups. As the treatment is uncorrelated with individual attributes, the first stage is in expectation always equal to 4. Under this research design, it is perfectly possible to study peer effects like in any standard 2SLS model described by equations 1 and 2.

If we were to run the same example under a typical design of random assignment to groups, we would still have twelve individuals who are randomly assigned to three groups of four. In this case, however, all groups would have the same expected value of peer attributes, $\overline{x}_g = \frac{1+12}{2} = 6.5$. This illustrates what Angrist (2014) describes as the weak variation problem. The only variation in the expectation across individuals comes from the exclusion bias. In particular, as individual 1 cannot be her own peer, the leave-out mean of the three students she can be paired with ranges from 2 to 12, with an average $\overline{x}_{g,-1} = \frac{2+12}{2} = 7$. On the other extreme, we have individual 12, with a leave-out mean of the three students she can be paired with that ranges from 1 to 11, with an average $\overline{x}_{g,-12} = \frac{1+11}{2} = 6$. The higher the level of attributes, the lower the peer leave-out mean. This negative correlation between individual attributes and the peer leave-out mean signals there is exclusion bias in random assignments to groups.

D Psychological Tests

This section describes in detail the psychological tests that were used to construct the sociability index.

In addition to the Big Five personality traits and the peers' perceptions measures described in section 5.1, the tests used to construct the sociability index are:

D.1 The Big Five

The most widely accepted taxonomy of psychological traits, both in the literature and in my data, is the Big Five (McCrae and John, 1992; John and Srivastava, 1999).[24] The *American Psychology Association Dictionary* defines the Big Five personality traits as follows (Table 1.1 in Almlund et al. (2011)):

1. Conscientiousness: the tendency to be organized, responsible, and hardworking.

2. Openness to Experience: the tendency to be open to new aesthetic, cultural, or intellectual experiences.

3. Extraversion: an orientation of one's interests and energies toward the outer world of people and things rather than the inner world of subjective experience; characterized by positive affect and sociability.

4. Agreeableness: the tendency to act in a cooperative, unselfish manner.

5. Neuroticism or Emotional Stability: Emotional Stability is "predictability and consistency in emotional reactions, with absence of rapid mood changes." Neuroticism is a chronic level of emotional instability and proneness to psychological distress.

Two traits from the Big Five are linked to social skills: extraversion[25] agreeableness[26]. Empirical evidence shows that extraversion is associated with good labor market outcomes (Fletcher, 2013), and that agreeableness influences occupational decisions (Almlund et al., 2011; Cobb-Clark and Tan, 2011). These results are consistent with a study by Deming (2017) that concludes that the labor market increasingly rewards social skills. I also include openness to experience[27] in the index as previous research shows that it is associated with leadership (Nieb and Zacher, 2015; Özbağ, 2016; Javed et al., 2020). In the COAR Network, it is also the trait with the largest predictive power on the number of peers that identify a student as a leader. The results are robust to excluding openness to experience from the index.

D.2 Altruism

The altruism self-reported scale was developed by Rushton et al. (1981). The test used in the COAR network is composed of 17 items. The score on the test is found to predict criteria such as peer ratings of altruism, completing an organ donor card, and paper-and-pencil measures of prosocial orientation (Rushton et al., 1981). More recent evidence shows that the score on the test is related to spontaneous smiles—which is an important signal in the formation and maintenance of cooperative relationships (Mehu et al., 2007). Likewise, there is evidence that the score on the test is related to charity giving but not to blood donor donation behavior (Otto and Bolle, 2011).

[24] Almlund et al. (2011) summarizes the Big Five personality traits and their application to economics. Likewise, Akee et al. (2018); Donato et al. (2017); Kranton and Sanders (2017) provide recent evidence of the Big Five in economics research.

[25] The facets of extraversion are: warmth (friendly), gregariousness (sociable), assertiveness (self-confident), activity (energetic), excitement seeking (adventurous), and positive emotions (enthusiastic).

[26] The facets of agreeableness are: trust (forgiving), straight-forwardness (not demanding), altruism (warm), compliance (not stubborn), modesty (not show-off), tender-mindedness (sympathetic).

[27] Openness involves six facets or dimensions, including active imagination (fantasy), aesthetic sensitivity, attentiveness to inner feelings, preference for variety, and intellectual curiosity.

D.3 Leadership

The leadership scale corresponds to the leader behavior questionnaire developed in Spanish by Castro-Solano (2007). It is based on the theory of Yukl (2013). The scale measures three components of leadership: (1) behaviors guided towards tasks, (2) behaviors guided towards others, and (3) behaviors guided towards changes. In my data, there is a positive correlation between the score on the scale and the number of peers who perceived the student as a leader.

D.4 Empathy

The empathy scale corresponds to the Basic Empathy Scale developed by Jolliffe and Farrington (2006). The scale is composed of two factors: cognitive and emotional empathy. The scale has been validated in other contexts: when applied to adults (Carre et al., 2013) and the Spanish version (Villadangos et al., 2016). It has also been affirmed that students who report higher scores in socially aversive personalities (psychopathy, narcissism, and Machiavellianism) have a low score on the scale (Wai and Tiliopoulos, 2012). Likewise, Gambin and Sharp (2018) show that a low score on the test is associated with guilt and depressive symptoms.

D.5 Intercultural Sensitivity

This 24-item scale of intercultural sensitivity was developed by Chen and Starosta (2000). The authors define intercultural sensitivity as: "*a person's ability to develop a positive emotion towards understanding and appreciating cultural differences that promotes appropriate and effective behavior in intercultural communication.*" The scale comprises two factors: positive and negative reactions to intercultural interactions. Evidence shows that there is a positive correlation between intercultural sensitivity and compassion in nurses (Arli and Bakan, 2018), that American student scores depend on religious affiliation and the number of times they have traveled outside the US (Gordon and Mwavita, 2018), and that Iranian university students have demonstrated a strong relationship between intercultural sensitivity and ethnic background.

D.6 Emotional Intelligence

Emotional intelligence is defined as individuals' ability to recognize their own emotions and those of others, discern between different feelings and label them appropriately, use emotional information to guide thinking and behavior, and manage and/or adjust emotions to adapt to environments or achieve one's goal(s) (Colman, 2009). The emotional intelligence test corresponds to the scale developed by Law et al. (2004). The test comprises 16 items and has four factors: self-emotional appraisal, uses of emotion, regulation of emotion, and others' emotional appraisal.

D.7 The Reading the Mind in the Eyes Test

This test aims to assess how well people can read others' emotions just by looking at pictures of their eyes. It is a multiple-choice test with 36 items. For each item, the respondent has to identify the corresponding emotion expressed in a pair of eyes; four choices are given for each question. According to Deming (2017), this test is a reliable measure of social skills since it relates to social value orientation (Declerck and Bogaert, 2008), a social intelligence factor, and performance in groups (Woolley et al., 2010), and individual teamwork abilities (Weidmann and Deming, 2020).

D.8 Achievement Goals

While not part of the construction of the social skills index, students completed the *Achievement Goal Questionnaire* (J. Elliot and Murayama, 2008). Achievement goals are conceptualized as cognitive–dynamic aims that focus on competence. The test comprises 12 items and has four factors: mastery approach goal items, mastery avoidance goal items, performance-approach goal items, and performance-avoidance goal items. The last two items are related to goals in comparison with peers and are the ones I use as part of self-confidence in academic skills.

E Theoretical Framework for the Role of Beliefs

In this section, I present a simple theoretical framework to understand how the formation of beliefs about abilities can drive peer effects. Overall, there are two mechanisms for self-confidence to improve students' outcomes. First, if ability and effort are complements in the education production function, students with higher confidence will exert more effort (Benabou and Tirole, 2002).

To illustrate this, let's consider the following education production function that depends on effort e_i and ability a_i.

$$y_i = a_i + \theta e_i + \gamma a_i e_i, \tag{E.1}$$

with $\theta > 0$ (effort improves the output) and $\gamma > 0$ (effort and ability are complements). The utility of student i is $u_i = y_i - \frac{c}{2}e_i^2$, where c parametrizes the marginal cost of effort. The optimal effort level of the student would be given by: $e_i^* = \frac{\theta+\gamma a_i}{c}$. When students have imperfect information then students take expectation over the ability distribution, such that:

$$e_i^* = \frac{\theta + \gamma \mathbb{E}\left[a_i\right]}{c}.$$

Hence, two students with the same level of ability (a_i) but different beliefs ($\mathbb{E}\left[a_i\right]$) would have different outcomes. By having higher self-confidence, students are incentivized to exert more effort, and this can improve their performance.

The second mechanism for self-confidence to affect performance is a direct one. Compte and Postlewaite (2004) introduce a model that explains how a person's psychological state can affect performance. In their model, the probability of success depends on a person's level of confidence, captured by her perception of success in previous cases. For example, a student who is more confident about her chances of making friends is more likely to make these friendships, and a student who is more confident in her math skills would have a higher score on a test.

A simple way of introducing the direct effect into the education production function is by including a parameter of self-confidence, $\kappa(\cdot)$ in equation E.1:

$$y_i = \kappa\left(\mathbb{E}\left[a_i\right]\right)\left(a_i + \theta e_i\right), \tag{E.2}$$

with $0 \leq \kappa(\cdot) \leq 1$, and $\kappa'(\cdot) > 0$. Notice that in equation *E.2*, I set $\gamma = 0$. The idea behind this production function is that even without complementarity between effort and ability, higher self-confidence can increase output.