

Proximity Can Induce Diverse Friendships: A Large Randomized Classroom Experiment

JULIA M. ROHRER – TAMÁS KELLER – FELIX ELWERT

CERS-IE WP – 2020/53
December 2020

<https://www.mtaki.hu/wp-content/uploads/2020/12/CERSIEWP202053.pdf>

CERS-IE Working Papers are circulated to promote discussion and provoke comments, they have not been peer-reviewed.
Any references to discussion papers should clearly state that the paper is preliminary.
Materials published in this series may be subject to further publication.

ABSTRACT

Can outside interventions foster socio-culturally diverse friendships? We executed a large field experiment that randomized the seating charts of 182 primary-school classrooms (N=2,996 students) for the duration of one semester. We found that being seated next to each other increased the probability of a mutual friendship from 15% to 22% on average. Furthermore, induced proximity increased the latent propensity toward friendship equally for all students, regardless of students' dyadic similarity with respect to educational achievement, gender, and ethnicity. However, the probability of a manifest friendship increased more among similar than among dissimilar students. Our findings demonstrate that a scalable light-touch intervention can affect face-to-face networks and foster diverse friendships in groups that already know each other, but they also highlight that transgressing boundaries defined by ethnicity and gender remains an uphill battle.

JEL codes: C93, I21, I24, J18, Z13

Keywords: Friendship formation, Social networks, Diversity, Homophily, Randomized field experiment, Deskmates

Julia M. Rohrer

Department of Psychology, University of Leipzig & International Max Planck Research School on the Life Course (LIFE), Max Planck Institute for Human Development, Berlin

and

Tamás Keller

Computational Social Science - Research Center for Educational and Network Studies, Centre for Social Sciences, Budapest
Institute of Economics, Center for Economic and Regional Studies, Budapest
TÁRKI Social Research Institute, Budapest

and

Felix Elwert

Department of Sociology & Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison

Acknowledgment

This research is funded by a grant from the Hungarian National Research, Development and Innovation Office (NKFIH), Grant number: FK 125358 to Tamás Keller, and by a Vilas Faculty Mid-Career Award from the University of Wisconsin-Madison to Felix Elwert. Tamás Keller acknowledges support from János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

A közelség segíti a társadalmi-kulturális szempontból sokszínű barátságok kialakulását: Egy randomizált osztálytermi terepkísérlet

ROHRER, JULIA M. – KELLER TAMÁS – ELWERT, FELIX

ÖSSZEFOGLALÓ

Elősegíthető-e külső beavatkozásokkal a társadalmi-kulturális szempontból sokszínű barátságok kialakulása? A kérdés megválaszolására terepkísérletet végeztünk, melynek során 182 általános iskolai osztályban randomizáltuk az ültetési rendet, és azt egy egész féléven keresztül fenntartottunk (N = 2996 tanuló). Megállapítottuk, hogy az egymás mellett ülő padtársak esetében a kölcsönös barátság kialakulásának valószínűsége 15%-ról 22%-ra nőtt. A padtársi kapcsolattal indukált közelség egyformán növelte a barátság iránti látens hajlandóságot minden diák esetében, függetlenül attól, hogy az egymás mellett ülő padtársak neme vagy etnikuma azonos volt-e, illetve hogy tanulmányi teljesítményük hasonló volt-e. A megvalósult barátságok valószínűsége azonban jobban nőtt az egymáshoz hasonló, mint az egymástól különböző diákok esetében. Eredményink azt mutatják, hogy egy olyan könnyen felskálázható, egyszerűen megvalósítható beavatkozás, mint a padtársak egymás mellé ültetése, hatással lehet a diákok között kialakuló kapcsolati hálózatokra, és sokszínű barátságok kialakulásához vezethet, akkor is, ha a diákok már ismerik egymást. Eredményink ugyanakkor felhívják a figyelmet arra is, hogy az etnikai és nemi határvonalak átlépése egy ilyen egyszerű beavatkozás számára továbbra is kihívás marad.

JEL: C93, I21, I24, J18, Z13

Kulcsszavak: Barátság kialakulása, Társadalmi kapcsolatháló, Sokféleség, Homofília, Randomizált terepkísérlet, Padtársak

Running head: Proximity, Diversity, and Friendship

Proximity Can Induce Diverse Friendships: A Large Randomized Classroom Experiment

Julia M. Rohrer

Department of Psychology, University of Leipzig & International Max Planck Research School on the Life Course (LIFE), Max Planck Institute for Human Development, Berlin

Tamás Keller

Institute of Economics, Center for Economic and Regional Studies, Budapest
Computational Social Science - Research Center for Educational and Network Studies,

Hungarian Academy of Sciences, Budapest

TÁRKI Social Research Institute, Budapest

Felix Elwert

Department of Sociology & Department of Biostatistics and Medical Informatics, University
of Wisconsin-Madison

Author Note

All analysis scripts underlying this article can be found on the Open Science Framework,

https://osf.io/4vjc5/?view_only=92e873da2b834c9892d3925edda90f96.

Abstract

Can outside interventions foster socio-culturally diverse friendships? We executed a large field experiment that randomized the seating charts of 182 primary-school classrooms ($N = 2,996$ students) for the duration of one semester. We found that being seated next to each other increased the probability of a mutual friendship from 15% to 22% on average.

Furthermore, induced proximity increased the latent propensity toward friendship equally for all students, regardless of students' dyadic similarity with respect to educational achievement, gender, and ethnicity. However, the probability of a manifest friendship increased more among similar than among dissimilar students. Our findings demonstrate that a scalable light-touch intervention can affect face-to-face networks and foster diverse friendships in groups that already know each other, but they also highlight that transgressing boundaries defined by ethnicity and gender remains an uphill battle.

Keywords: Friendship formation, social networks, diversity, homophily, field experiment

Proximity Can Induce Diverse Friendships: A Large Randomized Classroom Experiment

Friendships matter because social networks shape outcomes ranging from health behaviours to criminal activity and socio-economic achievement (Burt & Peterson, 1998; Christakis & Fowler, 2009; Clark & Loheac, 2007; Granovetter, 1973; Granovetter, 2005; Rivera, Soderstrom, & Uzzi, 2010; Stadtfeld, Vörös, Elmer, Boda, & Raabe, 2019).

Friendship networks, however, are strongly constrained by homophily: Humans have the well-documented tendency to form and maintain relationships with those who resemble themselves along dimensions such as gender, race, ethnicity, age, religion, and education (e.g., Lazarsfeld & Merton 1954; McPherson, Smith-Lovin, & Cook, 2001). Thus, social networks often lack diversity, and their benefits are distributed unevenly.

Previous research suggests that proximity between individuals can lead to friendships. In the present study, we investigate to which extent proximity can lead to friendships that transgress group boundaries imposed by homophily. For this purpose, we conducted a large-scale field experiment that randomly seated Hungarian primary school students in 3rd to 8th-grade classrooms¹ next to each other. To our knowledge, this is the first randomized experiment of this kind outside of the elite context of college, and the first study to explicitly investigate whether the effects of proximity on friendship are modified by similarity along the lines of gender, educational achievement, and ethnicity.

What Do We Know About the Effects of Proximity?

Prior research on the causal effects of proximity on friendships has focused on college freshmen. Multiple studies exploited natural experiments, such as alphabetical seating (Byrne, 1961; Byrne & Buehler, 1955). Others directly randomized the assignment of college roommates (Boisjoly et al., 2006), or the seating chart during an introductory meeting of

¹ Hungarian primary education is equivalent to ISCED 1 and ISCED 2 levels according to the international classification (elementary and middle school in the United States).

psychology freshmen (Back, Schmukle, & Egloff, 2008). Most of these studies find that proximity promotes friendship formation. Some studies further assessed downstream outcomes of induced proximity, establishing positive effects of being randomly assigned a Black roommate on White students' attitudes and behavior (Baker, Mayer, & Puller, 2011; Boisjoly, Duncan, Kremer, Levy, & Eccles, 2006; Camargo, Stinebrickner, & Stinebrickner, 2010; Carrell, Hoekstra, & West, 2016; Corno, La Ferrara, & Burns, 2019; Van Laar, Levin, Sinclair, & Sidanius, 2005; Marmaros & Sacerdote, 2006).

Open Questions

Prior findings are promising but leave two important questions open. First, prior work studied college freshmen, a highly selected and comparatively homogeneous elite in unusual circumstances, especially if they live in dorms.² Having recently relocated from the parental home to college, college students must quickly construct a new social network among strangers from scratch (Cutrona, 1982) and thus may be particularly susceptible to the effects of proximity. This raises the question whether previous findings generalize to other populations in more quotidian and scalable settings. The only recent (non-randomized) study on the effects of being seated next to each other on friendships among school-age children (van den Berg & Cillessen, 2015; van den Berg, Segers, & Cillessen, 2012) failed to find an effect of proximity on friendship nominations.

Second, little is known about the boundary conditions of the effects of proximity. Previous studies have addressed whether or not proximity can lead to inter-ethnic friendships (in the college context), but have neglected other potentially relevant dimensions of socio-cultural diversity. For example, it is unknown whether proximity can lead to friendships

²Only 8.3% of 18 to 24-year olds in the United States reside in student housing (U.S. Census Bureau, 2013-2017 American Community Survey 5-Year Estimates).

among students with different levels of educational achievement, which may expose lower-achieving students to positive role models; or whether proximity can lead to mixed-gender friendships, which may discourage the development of gendered attitudes and communication styles that are linked to power asymmetries in adulthood (Leaper, 1994; Maccoby, 1998; Leaper & Ayres, 2007; Mehta & Strough, 2009);

The Present Study

To investigate the effects of proximity on friendship in general, and the extent to which proximity can promote boundary-crossing friendships among school-age children in particular, we conducted a large pre-registered field experiment. We randomized the seating charts within 182 3rd through 8th grade classrooms in 40 primary schools in rural Hungary for the duration of the Fall 2017 semester (5 months). We assessed best-friend nominations at the beginning of the subsequent Spring 2018 semester to test the expectation that being seated next to each other had a positive causal effect on friendships (deskmate hypothesis). In contrast to previous studies among college students, we thus investigated the effects of proximity in a scalable environment (because nearly all children must attend school), at younger ages, and in groups that already know each other well (from 1st grade onwards).

Since humans tend to form friendships with self-similar others, the friendship-inducing effect of proximity is likely tampered by homophily. We therefore expect that induced proximity should promote friendship especially among individuals who resemble each other (modification-by-similarity hypothesis). In this study, we investigate effect modification by students' dyadic similarity along the three salient dimensions of gender, educational achievement, and ethnicity (Roma/non-Roma) to quantify the extent to which proximity can promote diverse friendships with respect to these categories.

Results

Deskmate Hypothesis: Effect of the Intervention on Friendships

We analyzed the effect of being seated next to each other (for the duration of one semester) on students' reciprocated friendships within the classroom (after the end of the semester) using Bayesian multi-membership multilevel probit models. We report results first as probit coefficients for the effects on students' latent continuous propensity toward friendship, and second as average marginal effects (AME) for the effects on students' predicted probability of a manifest reciprocated friendship. Since AMEs are non-linear functions of the probit coefficients, they answer different questions and may lead to qualitatively different conclusions (see Methods).

The results show that sitting next to each other had a large positive effect on students' friendships. The intervention increased the latent continuous friendship propensity of a dyad by an estimated $b_{\text{Deskmate}} = 0.27$ on the probit scale (Bayesian 95% credible interval CI_{95} : [0.19, 0.35]; $N_{\text{Students}} = 2,996$, $N_{\text{Dyads}} = 24,962$), and it increased the predicted probability of a manifest friendship by 7.0 percentage points (CI_{95} : [4.6; 9.4]), from 15.3 percent to 22.3 percent. This evidence confirms the deskmate hypothesis: induced proximity fostered friendships.

Our conclusions remained unchanged when including students in the analysis who were assigned to sit alone, and when analyzing directed friendship nominations (regardless of reciprocation). We found some variability of the deskmate effects across classrooms, but the differences were substantively small ($SD_{\text{Deskmate effect}} = 0.09$, CI_{95} : [0.00, 0.23]). More details of these additional analyses are reported on the Open Science framework.

Modification-by-Similarity Hypothesis: Moderating Role of Overall Similarity

We documented observational homophily by inspecting the association between reciprocated friendship and an index for dyadic similarity between students (Gower, 1971), which included students' gender (boy vs. girl), educational achievement (baseline grade-point average [GPA]), and ethnicity (Roma vs. non-Roma). As expected, there was a strong association between dyadic similarity and dyads' tendency to form a reciprocated friendship ($b_{Similarity} = 0.83$ per SD of the similarity index, $CI_{95}: [0.80, 0.86]$; $AME = 9.0$ percentage points, $CI_{95}: [8.5, 9.5]$, from low [$-1 SD$] to average similarity; $AME = 20.03$ percentage points, $CI_{95}: [19.4, 21.4]$, from average to high [$+1 SD$] similarity; $N_{Students} = 2,996$, $N_{Dyads} = 24,962$).

We tested the modification-by-similarity hypothesis by asking whether the causal deskmate effect varied with students' dyadic similarity. Support for the modification-by-similarity hypothesis was scale dependent, as is often the case in non-linear probability models (see Methods). We did not find evidence that similarity modified the deskmate effect with respect to the latent continuous friendship propensity, as the CI_{95} for the probit coefficient on the interaction term overlapped with zero ($b_{Deskmate*Similarity} = 0.07$, $CI_{95}: [-0.05, 0.18]$), and model fit did not improve when including it ($Diff_{ELPD} = 1.2 = 0.7 * SE_{Diff_{ELPD}}$ in favor of the more parsimonious model without the interaction term).

At the same time, students' dyadic similarity positively modified the deskmate effect with respect to the probability of manifest friendships (Figure 1, Panel A and B). The AME of sitting next to each other on the probability of manifest friendships was $AME_{Low} = 1.5$ percentage points ($CI_{95}: [0.2, 3.1]$) for dyads of low similarity; $AME_{Average} = 5.7$ percentage points ($CI_{95}: [3.4, 8.0]$) for dyads of average similarity; and $AME_{High} = 11.7$ percentage points ($CI_{95}: [7.7, 15.6]$) for dyads with high similarity. The 95% credible

intervals for the differences between the deskmate effects for dyads with low, average, and high similarity, respectively, comfortably excluded zero ($\Delta AME_{Average-Low} = 4.2$ percentage points, $CI_{95}: [0.3, 5.7]$; $\Delta AME_{High-Average} = 6.1$ percentage points, $CI_{95}: [2.6, 9.8]$).

Together, these results indicate that sitting next to each other equally increased the latent continuous friendship propensity for all dyads, and it also increased the probability of manifest friendships, even for fairly dissimilar dyads. But since the baseline friendship propensity was much larger among similar dyads (due to homophily), increasing the friendship propensity by a fixed amount pushed more similar dyads than dissimilar dyads across the threshold of manifest friendship. Thus, seating similar students next to each other resulted in more additional reciprocated friendships than did seating dissimilar students next to each other.

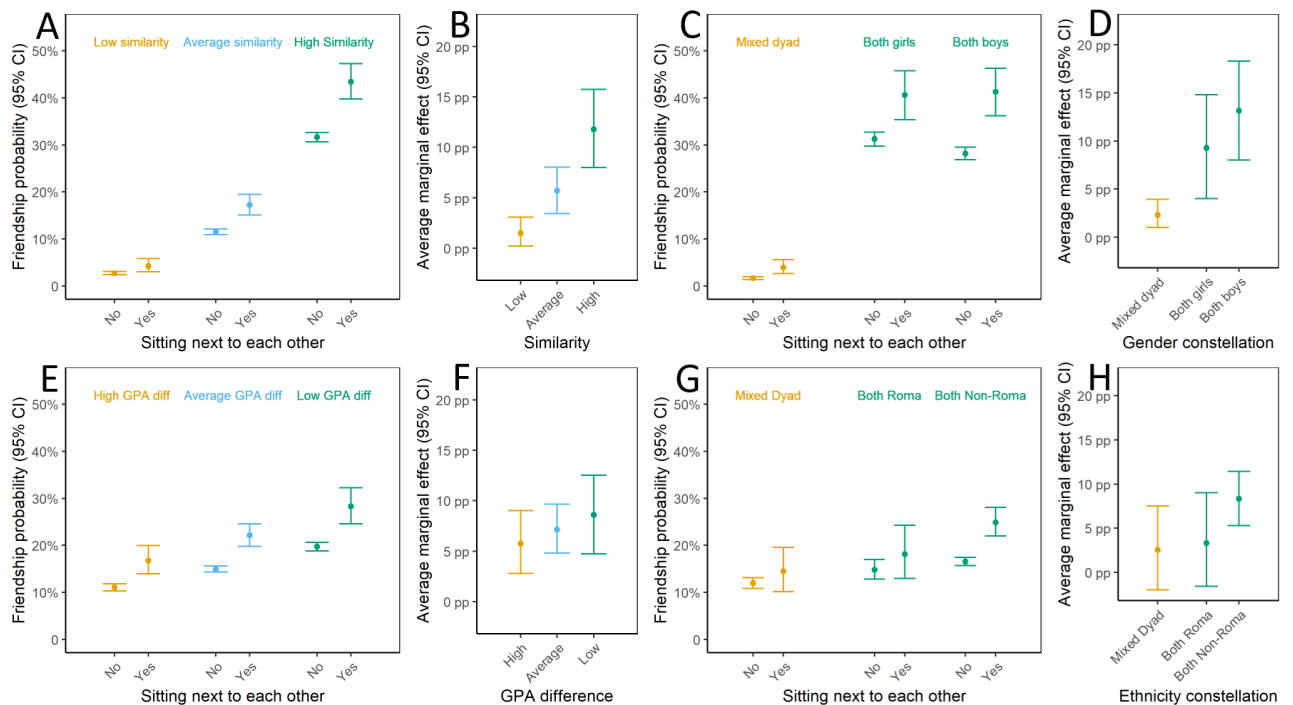


Figure 1. Model predictions for the effect of sitting next to each other on reciprocated friendships. Left column displays the predicted friendship probabilities (predictive margins); right column displays the corresponding average marginal deskmate effects (differences in the predicted friendship probabilities) in percentage points. Panel A and B: effect modification by overall similarity (low: -1 SD, high: +1 SD on Gower's similarity index based on gender, ethnicity, and baseline GPA). Panel C and D: effect modification by the gender composition of the dyad. Panel E and F: effect modification by absolute GPA difference between the students (low: -1 SD, high: +1 SD relative to the mean absolute difference). Panel G and H: effect modification by the ethnic (Roma/non-Roma) composition of the dyad.

Modification-by-Similarity Hypothesis: Moderating Influence of Gender, Educational Achievement and Ethnicity

To better understand the modifying role of dyadic similarity for the effect of proximity on friendship, we performed separate follow-up analyses along each dimension of similarity. A more detailed summary of the results, including all estimates, credible intervals, and model comparisons can be found on the Open Science Framework.

Gender. Results closely mirrored the results for the overall similarity index. There was strong associational homophily: Same-gender dyads ($N_{both\ boys} = 6,700$, $N_{both\ girls} = 5,848$) were much more likely to report a reciprocated friendship than mixed-gender dyads ($N_{mixed\ gender} = 12,414$). Evidence for the modification-by-similarity hypothesis was again scale dependent, with no modification of the effect of induced proximity on the latent continuous friendship propensity, but clear differences in the effects of sitting next to each other on the probability of manifest friendships. This can be seen in Figure 1, Panel C and D. Being seated next to each other increased the probability of a friendship among mixed-gender dyads by

$AME_{mixed\ gender} = 2.3$ percentage points ($CI_{95}: [1.0, 3.9]$), among all-female dyads by

$AME_{both\ girls} = 9.3$ percentage points ($CI_{95}: [4.0, 14.8]$); and among all-male dyads by

$AME_{both\ boys} = 13.1$ percentage points ($CI_{95}: [8.0, 18.3]$).

Educational achievement. There was once again clear evidence for associational homophily: the larger the absolute difference in baseline GPA between two students (controlling for dyad's mean baseline GPA), the smaller their propensity to report a reciprocated friendship. With respect to the modification-by-similarity hypothesis, we again found no effect modification with respect to the latent continuous friendship propensity. Furthermore, the intervention increased the probability of a manifest friendship for dyads in which students had similar or dissimilar grades (Figure 1, Panel E). While the AMEs of sitting next to each other on the probability of a manifest friendship increased slightly with the similarity of students' grades (Figure 1, Panel F), the 95% credible intervals for comparisons between the AMEs computed at different levels of dyadic similarity included zero.

Ethnicity. Again, we observed associational homophily: ethnically-matched dyads ($N_{both\ Non-Roma} = 16,811$, $N_{both\ Roma} = 2,851$) had a higher latent propensity for reciprocated friendships than dyads of mixed ethnicity ($N_{Mixed\ ethnicity\ dyad} = 3,932$). And again, we found no evidence that the ethnic constellation of the dyad modified the effect of sitting next to each other on the latent continuous friendship propensity. Considering the effects on manifest friendships, we found ambiguous support for effect modification by ethnic match. There was some rather weak evidence that the average marginal effect was higher for non-Roma dyads than for mixed ethnicity dyads ($AME_{Mixed\ ethnicity\ dyad} - AME_{both\ Non-Roma} = 5.8$ percentage points, $CI_{95}: [0.1, 11.2]$; see Figure 1, Panel G and H, although the upper bound of the CI_{95} crossed zero in two alternative model specifications, reported on the Open

Science Framework). There was no evidence that the AME for Roma dyads was higher than the AME for mixed-ethnicity dyads ($AME_{both\ Roma} - AME_{Mixed\ ethnicity\ dyad} = 0.8$ percentage points, $CI_{95}: [-6.4, 8.2]$). Taken together, this is at best ambiguous evidence for the modification-by-similarity hypothesis with respect to ethnicity, which would predict that the deskmate intervention is more effective at promoting friendships among *both* non-Roma and Roma dyads compared to dyads of mixed ethnicity.

Discussion

We executed a large pre-registered field experiment that randomized the seating charts in 182 3rd- through 8th-grade classrooms. We found clear evidence for a positive causal effect of proximity on friendship: sitting next to each other at the beginning of the semester substantially increased the probability of students' mutual best-friendship nominations after the semester had ended. This reverts the Null finding of the only recent, non-randomized, proximity intervention among school-age children (van den Berg & Cillessen 2015; van den Berg, Segers, & Cillessen, 2012).

Crucially, our study contributes nuanced new findings regarding the interactions between proximity and homophily in friendship formation. First replicating prior findings about the importance of homophily as a descriptive characteristic of friendship networks, we established that friendships were more likely to occur between students who shared the same gender, similar levels of academic achievement, and the same ethnicity. Next, we newly investigated the extent to which similarity modified the causal effect of being seated next to each other. Encouragingly, we found no evidence that induced proximity affected the latent continuous propensity towards friendship differentially for similar or dissimilar dyads of students. But since the effect of a given increase in the latent propensity toward friendship on the formation of a manifest friendship also depends on the dyad's baseline propensity toward

friendship, and since more similar dyads have a greater baseline propensity toward friendship (homophily), the intervention was more successful at inducing manifest friendships among similar students than among dissimilar students.

Induced spatial proximity nevertheless succeeded in inducing diverse friendships. Randomly seating boys and girls next to each other doubled their probability of nominating each other as best friends (from less than 2 to 4 percentage points). The intervention also substantially increased friendships between students with strong and weak baseline GPAs (from 11 to 17 percentage points). Finally, seating Roma and non-Roma students next to each other increased friendships across ethnic lines (from 12 to 15 percentage points), although this estimate is beset with statistical uncertainty, due to relatively small numbers of Roma students in the sample.

Thus, we conclude that even small changes in spatial proximity can substantially affect friendships, not only among the strangers studied in previous research, but also in groups that already know each other well. This documents that friendship networks remain malleable long after intra-group friendships have presumably been established. Furthermore, proximity also increases the propensity towards friendships, and the probability of manifest friendships, that transgress socio-cultural group boundaries—even as the transformations of latent propensities into manifest friendships remains to some extent an uphill battle against pervasive homophily.

Re-seating students is a low-cost and scalable intervention. Friendships across gender, achievement, and ethnical divides formed at a young age likely contribute to the development of social skills and shape attitudes with lasting consequences. This suggests the exciting possibility that targeted, low cost, and scalable interventions may reshape social networks to foster positive life outcomes for students and improve inter-group relations.

Acknowledgments

This research is funded by a grant from the Hungarian National Research, Development and Innovation Office (NKFIH), Grant number: FK 125358, and by a Vilas Faculty Mid-Career Award from the University of Wisconsin-Madison to Felix Elwert. The support from the János Bolyai Research Scholarship of the Hungarian Academy of Sciences and from the New National Excellence Program (ÚNKP) of the Ministry of Human Capacities are acknowledged (Grant number: ÚNKP-19-4-BCE-07). We would like to thank Steffen Nestler, Felix Schönbrodt, Alexander J. Etz, Stefan C. Schmukle, Michael Sobel, Benjamin Rosche, and Jingying He for advice. All errors are ours.

Materials and Methods

Pre-registration. We follow a detailed pre-analysis plan, filed at the RCT registry of the American Economic Association before the receipt of any outcomes data on April 13, 2018.³ Deviations from the plan are explained in the Supplementary Material and on the project page on the Open Science framework (https://osf.io/4vjc5/?view_only=92e873da2b834c9892d3925edda90f96), which archives all analytic scripts and more detailed model summaries.

³ See <https://doi.org/10.1257/rct.2895-1.0>

Study Overview. We executed a large-scale field experiment in primary schools in Hungary (általános iskola). Classroom-seating charts were randomized for the duration of the fall semester (September 2017 through January 2018). Outcome variables were collected through student surveys between February and April 2018. The analytic sample for the central analyses consisted of $N = 2,996$ students (forming 24,962 dyads) within 182 3rd through 8th grade classrooms at 40 schools. Of these students, 48.2% ($N = 1,447$) were female; and 22.2% were of Roma ethnicity ($N = 666$). Ethnicity was missing for 4.5% of the sample ($N = 136$). Students' ages ranged from 8 to 17 years ($M = 11.88$, $SD = 1.80$); the high end of the age range is due to students who had to repeat classes. Class sizes ranged from 10 to 33 students ($M = 19.42$).

Recruitment and Sample. In the spring of 2017, we contacted all primary schools in 7 contiguous counties of central Hungary, excluding the capital city of Budapest, via the heads of the local school districts to elicit information about classroom layouts and seating practices. We aimed to enroll all 3rd-8th grade classrooms in which (1) teachers would implement our randomized seating chart in three core subjects: Hungarian literature, Hungarian grammar, and mathematics; (2) all students would receive instruction in these subjects together (e.g., no ability grouping); (3) the classroom layout would form a grid of freestanding forward-facing 2-person desks.

After obtaining initial participation agreements from principals and teachers at 55 schools and dropping schools and classrooms that did not meet our inclusion criteria (see pre-analysis plan for details), the pre-analysis plan anticipated a sample of $N = 3,814$ students across 195 classrooms at 41 schools. The pre-analysis plan also stipulated additional exclusion criteria going forward. Following these pre-registered criteria, we dropped (in this order) 13 classrooms (containing 226 students) in which fewer than 30% of students answered the friendship-nomination item; 391 students who did not answer the friendship-

nomination item; 36 students with missing values on at least four of the seven variables comprising the similarity index; and 113 students who were assigned to sit alone (as a robustness check, we also report models including those students). Subsequent data inspection resulted in the exclusion of 11 doubly entered students, 5 students whose classrooms were smaller than the pre-registered minimum size of 10, and 36 students who had left their classrooms before the intervention. The final analytic sample consisted of $N = 2,996$ students forming 24,962 dyads within 182 3rd through 8th grade classrooms at 40 schools.

Pre-registered balance checks indicate excellent balance on all baseline covariates in the final analytic sample (for details, see Open Science Framework).

Intervention and Exposure Variable. Before the start of the fall semester 2017, we randomly assigned the students within each classroom to freestanding forward-facing two-person desks via unconstrained random partitioning, using a random number generator. We based the randomization on the class rosters from the preceding spring semester and stipulated a replacement algorithm to account for the small number of student who would exit or enter the class roster during the summer.⁴ We call the seating chart resulting from randomization and algorithm-compliant replacements the “intended seating chart.” The intended seating chart underlies all of our analyses (intention-to-treat analyses). Our central experimental exposure variable is coded = 1 for each dyad within a classroom that comprises deskmates in the intended seating chart, and = 0 otherwise.

Teachers were instructed to employ the intended seating chart in three subjects—mathematics, Hungarian literature, and Hungarian grammar—from the first day of classes

⁴ We instructed teachers to fill the seats of exited students with entering students from left to right, front to back, in alphabetic order of entering students’ surnames. This replacement rule plausibly preserves randomization.

(September 1, 2017) for the duration of the fall semester (January 31, 2018). Teachers were permitted to reseat students, but were asked to preserve deskmate assignments wherever possible.⁵ We assessed compliance via teacher reports after the second week of classes and via classroom visits by our field-staff throughout the fall semester; 94.4 percent of the dyads in which students actually sat next to each after the second week of classes comprised students who were supposed to sit next to each in the intended seating chart.

Baseline Covariates and Similarity Index. At the beginning of the study, classroom teachers reported students' characteristics including student's gender (male vs. female); ethnicity (Roma Hungarian vs. non-Roma Hungarian); and end-of-semester grades for spring 2017 in Hungarian literature, Hungarian grammar, mathematics, diligence and behavior. Grades ranged from 1 (worst) to 5 (best) for literature, grammar, and mathematics; and from 2 to 5 for diligence and behavior. From these five grades, we generated the grade point average (GPA). Following the pre-analysis plan, missing teacher reports of baseline covariates were filled in with students' self-reports collected at endline.

To quantify the similarity between students, we calculated Gower's (1971) general coefficient of similarity for each dyad within a classroom. Gower's index is a simple metric to quantify the similarity between two units along a number of variables that may be qualitative and/or quantitative. We calculated the similarity index based on students' gender, ethnicity, and baseline grades in literature, grammar, mathematics, behavior, and diligence. Gender and ethnicity were each weighted by a factor of 1; every baseline grade was weighted by a factor of 1/5 (i.e., all grades together received a weight of 1). Gower's index ranges from 0 (maximum possible dissimilarity along all dimensions) to 1 (perfect similarity along all dimensions). For example, a Roma girl and a non-Roma boy whose grades are all at opposing

⁵ For example, if a student had to be moved to the front of the classroom because of vision problems, we asked that her deskmate be moved with her.

ends of the scale would receive a similarity index of 0; two Roma girls with exactly the same grades would receive a similarity index of 1. We standardized the similarity index for further analysis.

Friendship nominations. At the end of the study, students were asked to nominate up to 5 of their “best friends” within the classroom as part of a written 45-minute in-class survey. In this study, the primary outcome is students’ reciprocated friendships, coded = 1 if both dyad members nominated each other as best friends and = 0 otherwise. As robustness checks, we also analyzed best-friend nominations within the classroom regardless of reciprocation.

Statistical Analyses

Deskmate hypothesis. We begin by evaluating the hypothesis that induced proximity fosters friendship between students. The deskmate effect is the net effect of sitting next to each other on friendship formation and friendship dissolution: students who were seated next to each other may either be induced newly to form a reciprocated tie, or to forego the dissolution of an existing tie.

Our primary specification is a Bayesian multi-membership multilevel probit model. The unit of analysis is the dyad (i.e. each row in the data represents one dyad of two students within a classroom who may or may not share a desk and who may or may not be friends). Following the conventional probit setup (Wooldridge, 2002), the model assumes that manifest (measured) binary friendship nominations occur when the dyad’s latent (unmeasured) continuous propensity for forming a reciprocated-friendship crosses a critical threshold; the latent friendship propensity is assumed to follow a normal distribution. We model this latent friendship propensity as a linear function of the binary deskmate indicator; a vector of classroom fixed effects to account for the experimental design that randomizes the

seating chart within classrooms; and random intercepts for the two students in each dyad to account for dependencies in the data as every student is member of multiple dyads.

We interpret the results of this model in two complementary ways to assess the causal effect of sharing a desk (i) on the latent continuous friendship propensity and (ii) on the probability of a manifest binary friendship nomination. First, following preferred practice in certain research fields (e.g., psychology), we present the probit coefficient for the deskmate indicator, which estimates the effect of sharing vs. not sharing a desk on the latent propensity for forming a reciprocated friendship. As always in probit (or logit) models, this parameter is identified only up to scale (Long, 1997), so that the coefficient should only be interpreted qualitatively as evidence about the direction of the deskmate effect on the latent friendship propensity.

Second, following preferred practice in other fields (e.g. economics and sociology), we present the average marginal effect (AME) of sharing vs. not sharing a desk on the probability of forming a manifest reciprocated friendship. Since probit models are non-linear probability models, the effect of deskmate exposures on the probability of manifest friendship nominations likely varies across dyads. The AME simply is the average of the effects of sitting vs. not sitting next to each other on the probability of forming a reciprocated friendship across all dyads. For clarity, we present AMEs alongside the average of the predicted probabilities (predictive margins) for forming a reciprocated friendship if all dyads were deskmates and if all dyads were not deskmates.

We evaluate the evidence for the deskmate hypothesis by computing Bayesian 95% credible intervals (CI_{95}) around our point estimates (probit coefficients, AMEs, and predictive margins, respectively). Bayesian CI_{95} are defined as the intervals into which the unobserved parameter falls with 95% probability, incorporating information from the

Bayesian priors. While there are fundamental differences between frequentist and Bayesian statistics, from a pragmatic perspective, Bayesian credible intervals and frequentist confidence intervals often lead to very similar numerical results (Albers, Kiers, & van Ravenzwaaij, 2018). We interpret credible intervals that exclude zero as evidence for the presence of an effect.

Importantly, since seating charts were randomized within classrooms, the estimated effect of sharing a desk—evaluated either as the effect on the latent friendship propensity or as the AME on the probability of manifest friendships—has a causal interpretation after controlling for classroom fixed effects.

Modification-by-Similarity hypothesis: Next, we evaluate the hypothesis that the causal effect of sharing a desk on friendship formation increases when students resemble each other more with respect to baseline characteristics. Since dyadic similarity was not randomized, our analyses do not identify the causal effect of similarity on friendship. Instead, we estimate (i) observational homophily, i.e. the extent to which similar individuals happen to befriend each other regardless of sharing a desk, and (ii) effect modification, i.e. the extent to which the causal deskmate effect varies with observed dyadic similarity (see VanderWeele 2015 on the difference between causal interaction and effect modification). The model evaluating observational homophily adds Gower's (1971) index of similarity to the model used to evaluate the deskmate effect. The model evaluating the similarity hypothesis (i.e., effect modification) further adds the product term of the index of similarity and the deskmate indicator. For graphical representations, we define “low” (“high”) values of Gower's similarity index as values falling 1 SD below (above) the sample mean across all dyads.

To explore which specific dimensions of the similarity index drive heterogeneity in the deskmate effect, we also estimate secondary models that allow the deskmate effect to

vary with indicators of the dyad's gender constellation (boy-boy, girl-girl, mixed-gender), ethnic constellation (both Roma, both non-Roma, mixed), or GPA. Analyses of modification by GPA incorporate two variables, the absolute GPA difference within the dyad and the dyad's mean GPA (to control for grade levels in order to isolate the role of grade differences). The analyses follow the same logic as for the index of similarity (i.e., we first add the indicator of similarity on the respective dimension to evaluate observational homophily, and then we further add its product term with the deskmate indicator to evaluate the modification-by-similarity hypothesis).

Special care must be taken when interpreting the coefficients of the product terms between the deskmate indicator and similarity measures for evidence about the modification-by-similarity hypothesis. Qualitative conclusions about interaction effects in non-linear models, such as probit or logit models, can depend on the scale of the outcome (Ai & Norton, 2003; Lotfus, 1978). For example, two variables that relate to the outcome additively on one scale may relate to the outcome multiplicatively after the model undergoes a non-linear transformation such that the scale of the outcome changes. Hence, when evaluating effect modification or interactions between variables in a probit model, it is possible that two variables (here, the deskmate indicator and dyadic similarity) statistically interact in their effect on the latent continuous outcome (here, the latent friendship propensity) but do not interact in their effect on the probability of the manifest binary outcome (here, the probability of friendship nominations), or vice versa—and this can result simply from transforming probit coefficients into AMEs *after* a given model has been estimated on given data.

Despite strong, and at times conflicting, preferences across methodological communities (e.g. Ai & Norton, 2003; Breen, Karlson, & Holm, 2018; Mize 2019; Simonsohn, 2017), no outcomes scale is inherently superior to another. Analysts who are interested in how the effect of sharing a desk on the latent friendship propensity is modified

by similarity would inspect the probit coefficient on the interaction between the deskmate indicator and the similarity measure. This would make sense, for example, for analysts who want to know whether all groups of students are similarly nudged toward more (or less) positive relations. By contrast, analysts who are interested in effect modification in the effect of sharing a desk on manifest friendship nominations would compare group-specific AMEs. This would make sense, for example, if analysts believe that sharp classifications into “friend” vs. “not a friend” matter for classroom dynamics. Since we are interested in both qualities of the friendship network (latent friendship propensities and manifest nominations), we present probit coefficients and AMEs (accompanied by the relevant predictive margins) alongside each other.

As before, we statistically evaluate our estimates (probit coefficients, AMEs, and predictive margins, respectively) by computing the relevant Bayesian CI_{95} . Additionally, in order to evaluate whether there is evidence for any variation between multiple groups (e.g. those defined by Gower’s index, or by gender, ethnicity, or GPA constellations), we compare models with and without deskmate-by-group interactions by inspecting the difference in their expected predictive accuracies, $Diff_{ELPD}$ (a Bayesian measure of model fit), computed via approximate leave-one-out cross-validation. Following convention, we conclude in favor of a model if $Diff_{ELPD}$ is at least twice its standard errors in its favor, $Diff_{ELPD} \geq 2 * SE_{Diff_{ELPD}}$.

Robustness checks: We explored several robustness checks for the deskmate and similarity hypotheses. First, in addition to reciprocated friendships, we also analyzed friendship nominations regardless of reciprocation. These analyses only differed from the previously described models in that (i) they included twice as many dyads, because every unordered dyad corresponds to two ordered sender-receiver dyads; and (ii) they included

random effects for senders and receivers. In contrast to the analysis of reciprocated friendships, lower-level units (dyads) were thus nested within one higher-level unit of the classification *sender* and one higher-level unit of the classification *receiver*, resulting in a cross-classified multilevel probit fixed effect model.

Second, we address methodological concerns (especially in economics) about probit fixed effects models, such as the models introduced above. First, it is known that non-linear fixed effects models can be problematic in small panels (here, small classrooms) (Greene, Han, & Schmidt, 2002). To address this concern, we re-estimated our primary models by substituting class-size indicators for the classroom fixed effect, thus replacing our pre-registered (and potentially problematic) fixed-effects model with a more conventional covariate-adjusted model. This substitution is permissible since the fixed effect is only needed to control for differences across classrooms in the probability that a given dyad is a deskmate dyad. Since this probability only depends on class size, controlling for class size is sufficient for causal identification. Second, to address the more general skepticism about non-linear models in parts of the social sciences, we estimated linear probability models (LPMs) for the probability of manifest friendship nominations. Among other advantages, LPMs, in contrast to probit and logit models, do not rest on distributional assumptions about the structural errors of the latent continuous friendship propensity implicit. These LPMs mirror the specification of the main analyses described above exactly, except for using a linear link function.

In short, results for all three types of models (probit with fixed effects, probit controlling for classroom size, LPM with fixed effects) were extremely similar. The largest discrepancy in the estimated AMEs across models was 1.6 percentage points, with most discrepancies well below 1 percentage points.

We present additional explanations behind all analyses, all model outputs, and a table contrasting the resulting estimates across different model specifications on the Open Science Framework.

Software: All models were estimated in the R (R Core Team, 2013) software package *brms* (Brückner, 2017, 2018) using R Studio (RStudio Team, 2015). We used the default Bayesian priors in *brms*, which are non-informative, or very weakly informative. All figures were created in *ggplot2* (Wickham, 2016).

IRB approval and consent

This study was reviewed and approved by the IRB offices at the Hungarian Academy of Science (data collection and analysis), and at the University of Wisconsin-Madison (data analysis). Consent was obtained at multiple points. First, we obtained consent from school administrators and teachers. Second, teachers obtained consent from students' parents.

References

- Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters*, *80*(1), 123–129.
- Albers, C. J., Kiers, H. A. L., & van Ravenzwaaij, D. (2018). Credible Confidence: A pragmatic view on the frequentist vs Bayesian debate. *Collabra: Psychology*, *4*(1). Retrieved from <https://collabra.org/articles/10.1525/collabra.149/>
- Back, M. D., Schmukle, S. C., & Egloff, B. (2008). Becoming friends by chance. *Psychological Science*, *19*(5), 439–440.
- Bigler, R. S. (1995). The Role of Classification Skill in Moderating Environmental Influences on Children's Gender Stereotyping: A Study of the Functional Use of Gender in the Classroom. *Child Development*, *66*(4), 1072.

- Breen, R., Karlson, K. B., & Holm, A. (2018). Interpreting and Understanding Logits, Probits, and Other Nonlinear Probability Models. *Annual Review of Sociology*, 44(1), 39–54.
- Bürkner, P. C. (2017). *brms*: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*. <https://doi.org/10.18637/jss.v080.i01>
- Bürkner, P. C. (2018). Advanced Bayesian Multilevel Modeling with the R Package *brms*. *The R Journal*. <https://doi.org/10.32614/RJ-2018-017>
- Burt, R. D., & Peterson, A. V., Jr. (1998). Smoking cessation among high school seniors. *Preventive Medicine*, 27(3), 319–327.
- Byrne, D. (1961). The Influence of Propinquity and Opportunities for Interaction on Classroom Relationships. *Human Relations; Studies towards the Integration of the Social Sciences*, 14(1), 63–69.
- Byrne, D., & Buehler, J. A. (1955). A note on the influence of propinquity upon acquaintanceships. *Journal of Abnormal and Social Psychology*, 51(1), 147.
- Christakis, N. A., & Fowler, J. H. (2009). *Connected: The surprising power of our social networks and how they shape our lives*. Little, Brown Spark.
- Clark, A. E., & Lohéac, Y. (2007). “It wasn’t me, it was them!” Social influence in risky behavior by adolescents. *Journal of Health Economics*, 26(4), 763–784.
- Gower, J. C. (1971). A General Coefficient of Similarity and Some of Its Properties. *Biometrics*, 27(4), 857–871.
- Granovetter, M. (1973). The Strength of Weak Ties. *The American Journal of Sociology*, 78(6), 1360–1380.
- Granovetter, M. (2005). The Impact of Social Structure on Economic Outcomes. *Journal of Economic Perspectives*, 19(1), 33–50.

- Greene, W., Han, C., & Schmidt, P. (2002). The bias of the fixed effects estimator in nonlinear models. Retrieved from <http://people.stern.nyu.edu/wgreene/nonlinearfixedeffects.pdf>
- Leaper, C. (1994). Exploring the consequences of gender segregation on social relationships. *New Directions for Child and Adolescent Development*, 1994(65), 67–86.
- Leaper, C., & Ayres, M. M. (2007). A meta-analytic review of gender variations in adults' language use: talkativeness, affiliative speech, and assertive speech. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, 11(4), 328–363.
- Long, J. S. (1997). Regression Models for Categorical and limited dependent variables. Sage: Thousand Oaks, CA.
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27(1), 415–444.
- Mize, T. D. (2019). Best practices for estimating, interpreting, and presenting nonlinear interaction effects. *Sociological Science*, 6, 81–117.
- R Core Team. (2013). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna*.
- Rivera, M. T., Soderstrom, S. B., & Uzzi, B. (2010). Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms. *Annual Review of Sociology*, 36(1), 91–115.
- RStudio Team. (2015). RStudio: Integrated development for R. *RStudio, Inc., Boston, MA*
URL <http://www.Rstudio.com>.
- Simonsohn, U. (2017). [57] Interactions in Logit Regressions: Why Positive May Mean Negative. Retrieved from Datacolada website: <http://datacolada.org/57>

- Stadtfeld, C., Vörös, A., Elmer, T., Boda, Z., & Raabe, I. J. (2019). Integration in emerging social networks explains academic failure and success. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(3), 792–797.
- van den Berg, Y. H. M., & Cillessen, A. H. N. (2015). Peer status and classroom seating arrangements: a social relations analysis. *Journal of Experimental Child Psychology*, *130*, 19–34.
- van den Berg, Y. H. M., Segers, E., & Cillessen, A. H. N. (2012). Changing peer perceptions and victimization through classroom arrangements: A field experiment. *Journal of Abnormal Child Psychology*, *40*(3), 403–412.
- VanderWeele, T. J. (2009). On the distinction between interaction and effect modification. *Epidemiology*, *20*(6), 863–871.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer.

Supplementary Materials

Supplementary Text

Deviations Between the Pre-Analysis Plan and the Reported Analyses

We followed the pre-specified data collection and data exclusion procedures. We used the pre-specified primary outcome (reciprocated best-friend nominations) and primary treatment variables (assigned desk mate at baseline and the pre-specified desk-mate similarity index) to test the pre-specified Hypothesis 1 (desk mate exposure): The probability of a best-friend relationship increases if j is i 's deskmate. The following sections explain the rationale behind deviations between the pre-analysis plan and the reported analysis. There were two major deviations: we re-specified our models as multilevel models, and we omitted the test of a second hypothesis that raised some unexpected complexities. In the last section, we summarize and justify minor deviations.

Multilevel Models Instead of Social Relations Regression Models.

In the pre-analysis plan, we specified that we were going to test these hypotheses in a hierarchical social relations probit regression model. We had to adapt this plan for multiple reasons.

First, we had planned to use the R package *amen* by Peter D. Hoff. However, it turned out that *amen* is not suited to accommodate the hierarchical structure of the data. Hence, we re-specified our models in *brms* as described in the main text. These models are very similar on a conceptual level, except for omitting the particular covariance structure not relevant for the present research question.

To check whether our *brms* specification results in similar results as the originally planned *amen* implementation, we ran both analyses for a random subset of 50 classrooms, applying the models to one classroom at a time. The point estimates of the deskmate effect were highly correlated, $r = .831$. While for most classrooms, models returned plausible estimates, some classrooms resulted in very large negative point estimates with considerable uncertainty (i.e., for 24% of the classrooms, effects were estimated to be smaller than -2). This has to do with the friendships within the respective classrooms: If *none* of the deskmate dyads reports a reciprocated friendship, the estimate would become negative infinity were it not for the regularizing influence of the prior. Because of differences in the model priors, these outliers (here defined as coefficients below -2) were much more extreme in *brms* (mean among outliers -234.01) than in *amen* (mean among outliers -4.08). When ignoring these outliers, the correlation between the estimates was almost perfect ($r = .992$) and the absolute differences in the deskmate probit coefficients between *brms* and *amen* were small ($M_{absolute\ difference} = 0.06$), indicating that both types of models lead to highly similar conclusions except for cases in which the data situation made it impossible to give a reliable estimate of the deskmate effect.

Notice that the negative outliers no longer appear in the combined overall analysis in *brms*, even when deskmate effects were allowed to vary by classroom (Figure S1): The nature of the multi-level model provides regularization, shrinking estimates in classrooms that do not allow for a reliable estimation of the effect towards the average effect.

Omission of Hypothesis 2

The pre-analysis plan specified another hypothesis, which has been omitted from this manuscript, Hypothesis 2 (similarity between class mate and desk mate): The probability of a best-friend relationship increases if classmate j resembles student i 's desk mate. There are

multiple complications with the model we had planned to test this hypothesis which eventually led us to exclude it from this manuscript.

First, somewhat trivially, the model in the pre-analysis plan was redundantly notated. The pre-specified analysis would have included (1) whether or not i and j were seated next to each other, (2) the similarity between i 's deskmate and student j (labelled "deskmate similarity"), and (3) the interaction of the two. But if i and j are deskmates, then the similarity between i 's deskmate (j) and student j will be 1 and the interaction term is 1. If the two students are not deskmates, the interaction term is 0. In other words, the interaction term is fully redundant with predictor (1). This issue is an inconsequential but unnecessary notational redundancy.

Furthermore, predictor (2), deskmate similarity, is actually causally affected by the deskmate variable (1). Hence, including (2) will distort the estimate of the effect of (1), the central treatment of interest. Therefore, even though we had planned to test both hypotheses within a single model, when testing Hypothesis 1, deskmate similarity ought not to be included or else the deskmate effect is no longer causally identified.

Apart from these issues, we discovered that even after correcting the model for the above noted endogeneity, the pre-registered model would falsely detect arbitrary "effects" in a simulated Null model. Fixing this issue will require methodological work that would go beyond the present manuscript. Since we regard the deskmate hypothesis as the main substantive point of interest in this study, we elected to reserve the other hypothesis for future work.

Smaller Deviations

Liking of Deskmate. The pre-analysis plan stated in the section Secondary Outcomes: "Second, we will analyze how much student I liked their desk mate, coded 1-5,

where 1 is best and 2 is least and ‘don’t know’ is coded as ‘neutral’ = 3 (Question 7).” While this variable has been collected, its utility to address any substantive questions is questionable, because this question has only been asked with respects to deskmates, not with respects to other students in the classroom—hence, it cannot be used as an alternative outcome of the deskmate intervention.

Missing values. The pre-analysis plan states that “If we have missing values on a covariate we will code the covariate as zero and include dummy variables controlling for missing status so that we do not lose observations.” This approach is controversial from a methodological perspective. Furthermore, the relevant variables for which we observe missing values (ethnicity, GPA) are never just used for the purpose of covariate adjustment but are instead central predictors of interest. Therefore, and because there were only few missing values (5.5% of dyads lacked information on ethnicity, 6.2% of dyads lacked information on GPA), we instead simply excluded dyads for which the relevant variables were missing.

Balance Checks

Following our preregistration plan, we tested for balance following Guryan, Kroft, and Notowidigdo (2009) by regressing each student’s own baseline characteristic (gender, ethnicity, GPA based on the five reported grades) on (1) the student’s deskmate’s baseline characteristic, (2) the leave-one-out mean characteristic in the classroom, and (3) classroom fixed effects. This procedure corrects for the artifactual correlation between students’ and their deskmates’ characteristics induced by randomly partitioning students to desks within classrooms (see e.g. Boozer & Cacciola, 2001). Standard errors were clustered at the school-level. Results indicated excellent balance, with no significant associations between students’ and deskmates’ GPAs ($p = .311$), gender ($p = .558$) or ethnicity ($p = .886$).

Additional Analyses

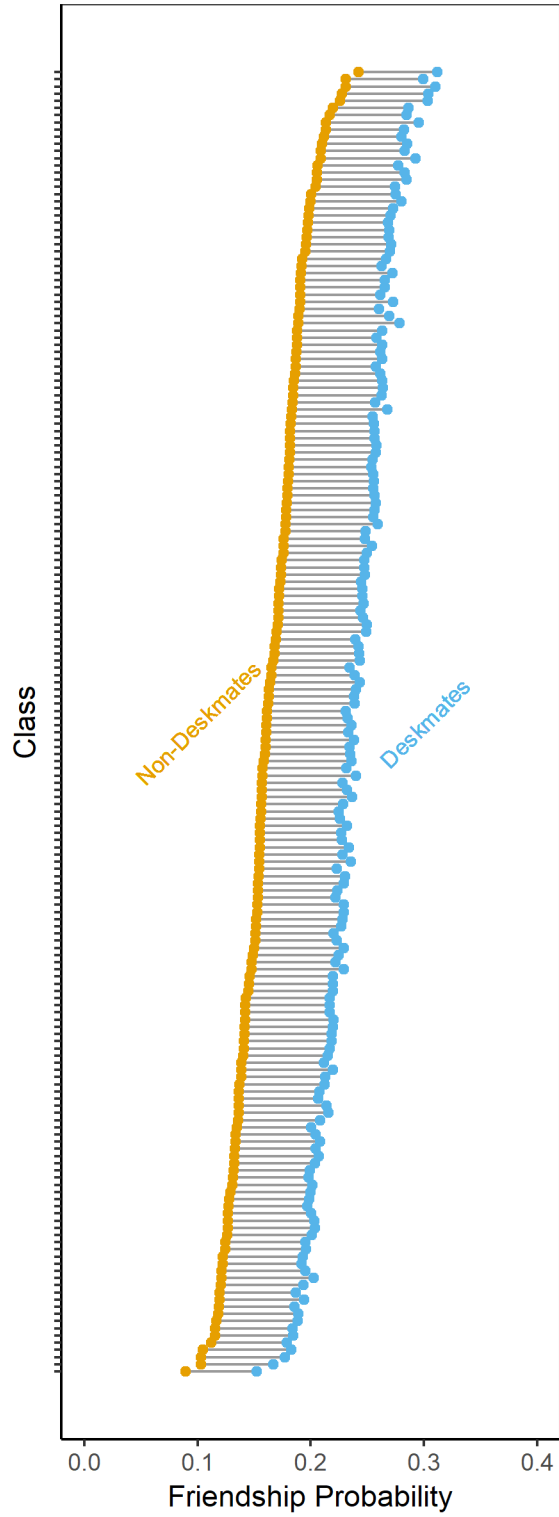
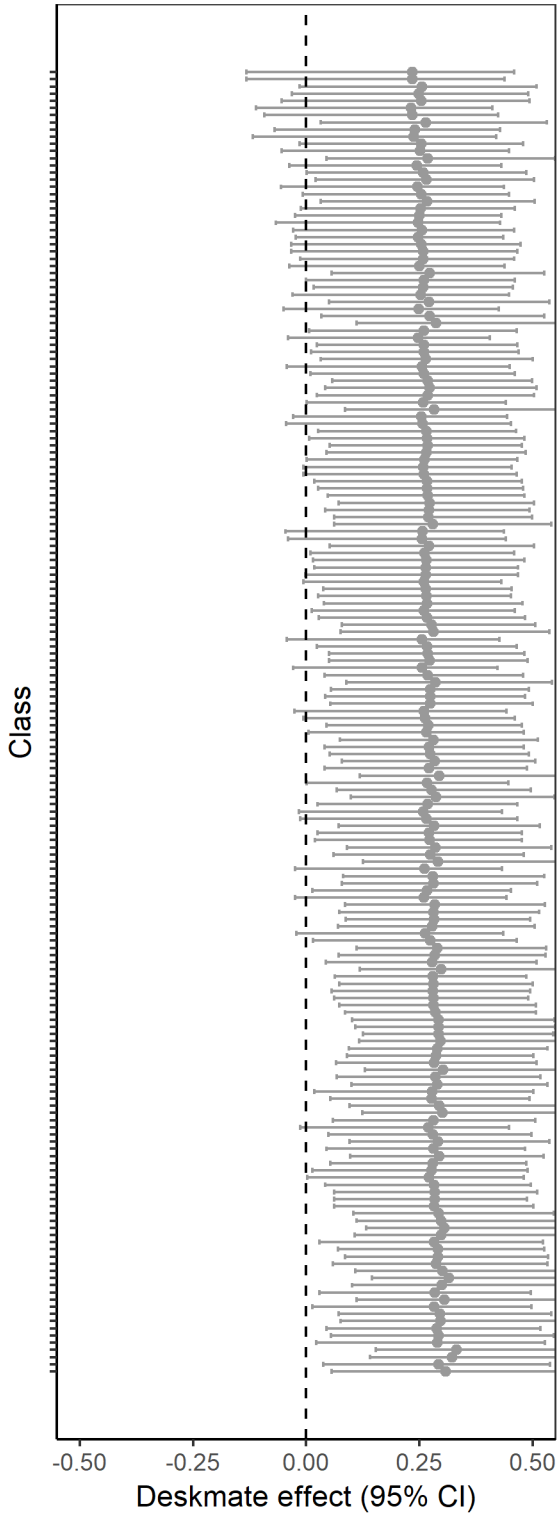


Figure S1. Heterogeneity of the deskmate effect across classrooms, probit coefficients from random effects model (left panel) as well as the corresponding model-implied friendship probabilities for non-deskmates versus deskmates (right panel). The difference between the each predicted probability for deskmates minus the predicted probability for non-deskmates is the classroom-specific AME.