



Feasible Peer Effects: Experimental Evidence for Deskmate Effects on Educational Achievement and Inequality

Tamás Keller,^{a,b,c} Felix Elwert^d

a) HUN-REN Centre for Social Sciences; b) HUN-REN Centre for Economic and Regional Studies; c) TÁRKI Social Research Institute; d) University of Wisconsin-Madison

Abstract: Schools routinely employ seating charts to influence educational outcomes. Dependable evidence for the causal effects of seating charts on students' achievement levels and inequality, however, is scarce. We executed a large pre-registered field experiment to estimate causal peer effects on students' test scores and grades by randomizing the seating charts of 195 classrooms (N=3,365 students). We found that neither sitting next to a deskmate with higher prior achievement nor sitting next to a female deskmate affected learning outcomes on average. However, we also found that sitting next to the highest-achieving deskmates improved the educational outcomes of the lowest-achieving students; and sitting next to the lowest-achieving deskmates lowered the educational outcomes of the highest-achieving students. Therefore, compared to random seating charts, achievement-discordant seating charts would decrease inequality; whereas achievement-concordant seating charts would increase inequality. We discuss policy implications.

Keywords: deskmates; peer effects; field experiment; achievement; inequality; Hungary

Citation: Keller, Tamás, and Felix Elwert. 2023. "Feasible Peer Effects: Experimental Evidence for Deskmate Effects on Educational Achievement and Inequality" *Sociological Science* 7: 806-829.

Received: July 12, 2022

Accepted: May 2, 2023

Published: November 6, 2023

Editor(s): Arnout van de Rijt, Peter Bearman

DOI: 10.15195/v10.a28

Copyright: © 2023 The Author(s). This open-access article has been published under a Creative Commons Attribution License, which allows unrestricted use, distribution and reproduction, in any form, as long as the original author and source have been credited.

COMPOSITIONAL peer effects—the causal effects of students' exposure to other students (Ogburn and VanderWeele 2014)—matter because they can influence educational choices and outcomes. Compositional peer effects have commanded attention ever since the Coleman Report (Coleman 1966, 1968) found a positive correlation between students' outcomes and the characteristics of their peers. Today, compositional peer effects inform research and policy on segregation, busing, school choice, selective admissions, affirmative action, ability tracking, and grouping. (for example, Bygren 2016; Hallinan 1994; Slavin 1990; Terrin and Triventi 2023).

Most past research in education has focused on compositional peer effects in large groups, such as schools and classrooms. This research, however, rarely isolates direct peer effects from competing mechanisms. For example, the total effect of educational tracking comprises direct peer effects, differential teacher quality and school resources across tracks, curricular differentiation, and the adjustment of teachers' expectations to their audience (Card and Giuliano 2016; Duflo, Dupas, and Kremer 2011; Hanushek and Wößmann 2006; Oosterbeek, Ruijs, and de Wolf 2023; Schiltz et al. 2019).¹ Studying smaller, sub-classroom environments benefits theory development because it helps isolate compositional peer effects from competing mechanisms. Deskmate effects, in particular, are a promising target of inference because although teachers may sort into schools and classrooms in response to school- or classroom level student achievement, they do not sort into classrooms in

response to the seating chart. Similarly, teachers are unlikely to adjust expectations, curricular demands, or teaching styles at the desk level. Hence, deskmate effects likely represent pure, direct, peer effects.

Studying micro-environmental peer effects, and deskmates in particular, may also benefit policy. Although macro-level interventions, such as school de-segregation and tracking, are hindered by cost and political controversy, intervening in seating charts is cheap and broadly accepted because teachers typically control the seating chart. Hence, even if micro-level peer interventions have small effects at the individual level, their feasibility facilitates scaling to generate large effects in the aggregate.

To study micro-environmental peer effects in education and their aggregate consequences (efficiency and inequality), we executed the largest randomized field experiment on the effect of close peers on educational outcomes by randomizing the seating charts in 195 classrooms across 41 Hungarian primary schools. Students were assigned to random deskmates for the duration of one semester. Following best-practice advice, we analyzed multiple learning outcomes, including teacher-assigned grades and machine-graded standardized tests; measured outcomes both during and after the intervention; pre-registered a detailed analysis plan to prevent p-hacking and hindsight bias; and we corrected our results for multiple-hypothesis testing to control the risk of “false discoveries,” that is, the over-interpretation of chance findings.

We present results for the effect of deskmates on students’ individual outcomes and results for the aggregate consequences of seating charts on mean achievement levels and inequality. Our primary confirmatory analyses found no dependable evidence that sitting next to a deskmate with higher prior achievement or sitting next to a female deskmate meaningfully affected learning outcomes for the average student. Exploratory analyses, however, revealed important effect heterogeneity. Although the middle half of students were not affected by sitting next to deskmates with the lowest or the highest baseline achievement, sitting next to the highest-achieving deskmates improved the educational outcomes of the lowest-achieving students, and sitting next to the lowest-achieving deskmates lowered the educational outcomes of the highest-achieving students. This implies that achievement-discordant seating charts, compared to random seating charts, would decrease inequality in educational outcomes, whereas achievement-concordant seating charts would increase inequality in educational outcomes—without affecting mean achievement in the classroom. For theory, our results highlight that even micro-level peer effects can affect distributional outcomes. For policy, these findings are promising because seating charts are easily manipulated in practice.

The article proceeds as follows: Section 2 lists theoretical mechanisms, states our expectations and reviews prior work. Section 3 describes the institutional setting. Section 4 details the experimental design. Section 5 explains estimation. Section 6 presents the results. Section 7 discusses the implications and concludes.

Expectations and Prior Research

Our theoretical expectations about the effects of deskmates on students' learning outcomes are informed by multiple related literatures on compositional peer effects in education, including the small literature on deskmate effects (Hong and Lee 2017; Keller and Takács 2019; Li et al. 2014; Lu and Anderson 2015; Wu, Zhang, and Wang 2023) and the sizeable literatures on dorm room assignments (for example, Foster 2006; McEwan and Soderberg 2006; Sacerdote 2001; Stinebrickner and Stinebrickner 2006), study groups (for example, Antecol, Eren, and Ozbeklik 2016; Booij, Leuven, and Oosterbeek 2016; Feld and Zölitz 2017), and educational tracking (for example, Card and Giuliano 2016; Duflo et al. 2011; Hanushek and Wößmann 2006; Oosterbeek et al. 2023; Schiltz et al. 2019).² Together, these literatures suggest that deskmate exposures may affect students' academic outcomes via multiple mechanisms. First, students may learn directly from their deskmates by asking them for help, guidance, or explanations (Hanushek et al. 2003). Second, students and deskmates may learn from each other by collaborating on in-class exercises and assignments (Lu and Anderson 2015). Third, students may benefit from a deskmate's quiet focus or be distracted by their disruptive behavior (Lazear 2001). Fourth, students may treat deskmates as role models and emulate their effort and study habits (Stinebrickner and Stinebrickner 2006). Fifth, students may illicitly copy their deskmate's written work, including assignments and exams, or otherwise cheat.³

Some of these potential mechanisms are specific to deskmate relationships, such as pairwise collaboration or copying on exams. Other mechanisms are merely especially salient for deskmates. For example, although disruptive students may affect the entire classroom, they will primarily distract their deskmates; and students are more likely to accept others as role models if they are friends, which proximity promotes (Rohrer, Keller, and Elwert 2021).⁴

Consequently, we pre-registered two primary hypotheses that sitting next to a deskmate with higher prior achievement (1), and sitting next to a girl (2), increases students' academic achievement at endline. These hypotheses are interrelated because deskmates' characteristics are interrelated, that is, girls, on average, are better students than boys.

Five prior studies investigated deskmate effects on students' academic outcomes, with differing results. Hong and Lee (2017) studied deskmate effects at an elite Korean university, using instrumental variables estimation. They found that sitting next to a deskmate with a one-standard-deviation higher midterm score increased students' own final exam score by 0.12 standard deviations. This is a sizeable estimate, especially because it represents a lower bound on the true deskmate effect under the authors' theoretical model. Point estimates are similar for men and women but only statistically significant for men. Sitting next to a higher-achieving deskmate was found to help the lowest and highest performing students, but not students in the middle of the achievement distribution.

Keller and Takács (2019) analyzed forty 10th-grade classrooms in seven underprivileged Hungarian high schools, using ordinary regression with classroom fixed effects. They, too, estimated large effects: a one-standard-deviation increase in

deskmates' 8th-grade reading scores increased students' 10th-grade reading by 0.12 standard deviations. No effect was found for mathematics test scores, and effects did not differ by ethnicity.

The usual concerns over observational peer effects estimation apply to the foregoing two studies (Angrist 2014; Bramoullé, Djebbari, and Fortin 2020; Elwert and Winship 2014; Shalizi and Thomas 2011), including the possibility of unobserved selection into deskmate allocations within classrooms (Keller and Takács 2019), possible violations of the instrument-variables exclusion restrictions (Hong and Lee 2017), and upward bias due to measurement error (Angrist 2014; Feld and Zölitz 2017).⁵

The three prior randomized experiments found much smaller, or no deskmate effects. Lu and Anderson (2015) randomized the seating chart in twelve 7th-grade classrooms in one Chinese middle school for the duration of one academic year. They found that a one-standard-deviation increase in deskmate's baseline test scores (aggregated across multiple subjects) increased students' endline test scores by only 0.02 standard deviations ($p=0.34$). Sitting next to a female deskmate, by contrast, increased students' test scores by 0.07 standard deviations ($p = 0.02$) on average, an effect that did not appear to differ between boys and girls and that the authors ascribe to an increase in cooperative learning.⁶ Li et al. (2014) and Wu et al. (2023) executed deskmate experiments in China with monetary incentives for peer performance. In sub-analyses without incentives, they found no evidence for causal effects of deskmates on achievement.

Prior research on deskmate effects has hardly explored implications for aggregate achievement, either for levels (efficiency) or inequality. Nonetheless, published results permit speculation.⁷ For example, it is easy to see that deskmate effects on aggregate achievement levels require effect heterogeneity—if all students benefited equally from every type of deskmate, then it would not matter, on average, where anybody sits. Hence, the absence of evidence for heterogeneity in the effect of sitting next to higher-achieving deskmates by gender (Hong and Lee 2017) and ethnicity (Keller and Takács 2019), just like Lu and Anderson's (2015) experimental null findings for effect heterogeneity at the desk level, imply no effect of seating charts on aggregate achievement levels one way or the other.

By contrast, when effect heterogeneity exists, it is possible that changing the seating chart may increase or decrease aggregate achievement levels. Predicting such consequences is difficult because they depend on the pattern and magnitude of the heterogeneity and on the extent to which various heterogeneity-relevant deskmate characteristics correlate with each other. Lu and Anderson's (2015) experimental finding for the effect of sitting next to girls suggests some leverage. Although the authors did not report evidence for differential effects of sitting next to a female deskmate on boys or girls, they did find that being fully surrounded by girls—front, back, and sides—increases girls' test scores by 0.2 standard deviations whereas being surrounded by girls may reduce boys' performance. Hence, gender segregation within the classroom as a whole (for example, boys seated on the left side of the room, girls on the right) or across classrooms may increase the mean achievement in the classroom.⁸

Effects of the seating chart on aggregate inequality do not require heterogeneous deskmate effects. For example, in gender-balanced classrooms where girls are better students than boys on average, Lu and Anderon's (2015) finding of equally positive effects of sitting next to a girl on both girls and boys implies that gender-concordant seating would increase inequality, whereas gender-discordant seating would decrease inequality. Similarly, Hong and Lee's (2017) finding that both high- and low-achieving students (but not middle students) benefit from sitting next to a high-achieving deskmate suggests that achievement-concordant seating might increase inequality.

In general, however, estimates from past research on deskmate effects, using relatively small samples, are too imprecise to offer a firm basis for speculation. Furthermore, whether these possible effects matter in practice is difficult to predict in the absence of exact calculations. We explore the consequences of our estimates for aggregate outcomes across multiple possible seating charts by simulation below.

Institutional Context

We study deskmate effects on students' learning outcomes in 3rd- to 8th-grade classrooms of rural Hungarian primary schools. Primary education in Hungary commences with first grade at age six and ends after 8th grade. Rural students attend the public school in their local catchment areas where school choice is negligible. Rural primary schools are not tracked by prior achievement or segregated by gender, ethnicity, or other criteria beyond the existing composition of the school's catchment area. Classrooms are quite stable so that students remain with the same classmates across grades.

Seating charts are typically determined by students' homeroom teachers. Teachers testify to their belief in the importance of seating charts for students' outcomes and often seat disruptive next to well-behaved students and lower-achieving next to higher-achieving students. Few teachers purposefully seat girls next to boys.⁹

Students spend a significant amount of time next to their deskmate, because seating charts rarely change and apply to all subjects taught in the same room.¹⁰ For example, 3rd to 8th grade students receive seven to ten 45-minute lessons per week in the three core subjects: Hungarian literature (reading), Hungarian grammar (writing), and mathematics in the same classroom, accounting for between one quarter and half of the average school day.¹¹ Because many subjects in rural schools are taught in the same room, the total exposure of students to their deskmates in our sample is likely substantially greater.

Deskmates have many opportunities to influence each other directly, for example, through collaborative learning. In a 2022 survey that we conducted among a nationwide sample of $N=656$ primary school teachers, most teachers reported that deskmates collaborate almost every lesson (61 percent) or at least once a week (95 percent). The three most common deskmate activities were helping each other to learn, working together, and developing social skills.

Students are graded on a combination of monthly written exams, end-of-semester exams, recitations, participation, and sometimes behavior.¹² Written work carries

the greatest weight for end-of-semester grades. Grades matter for students' grade advancement and admission to tracked upper-secondary schools in 9th grade.

Importantly, students are not graded "on a curve." Eighty-one percent of primary school teachers in our 2021 survey reported using absolute achievement cutoffs for assigning grades. Because grade distributions thus are not fixed, we can detect deskmate effects on average grades, should such effects exist.¹³

Experimental Design

Consent, Pre-registration and Research Transparency

The study was reviewed and approved by the IRB offices at the Center for Social Sciences, Budapest, and the University of Wisconsin-Madison. Consent was obtained from school districts, principals, teachers, and students' parents. We pre-registered coding choices and statistical analyses in a detailed pre-analysis plan (available at <https://www.socialsciceregistry.org/trials/2610>), which was submitted before receiving endline data. A replication package containing data, survey instruments, and analytic scripts is available at: <https://osf.io/ehjf8/>.

Recruitment, Random Assignment, and Compliance

In early 2017, we contacted all primary schools in seven contiguous rural counties of north-central Hungary via the heads of the local school districts. We obtained initial participation agreements from 55 schools, in which most 3rd-8th grade classrooms were anticipated to implement our randomized seating chart in at least three core subjects (Hungarian literature, grammar, and mathematics), and all students in a classroom would receive instruction in these subjects together (for example, no ability grouping).

Shortly before the start of the 2017 fall semester, we randomized students via unconstrained random partitioning to freestanding, two-person, and front-facing desks within each participating classroom. Desks were arranged on a grid and separated by aisles to ensure that each student had only one deskmate. We based randomization on the class rosters from the preceding spring semester and stipulated a replacement algorithm to account for changes to class rosters via exits and entries during the summer. We requested adherence to the random seating chart for the duration of the entire fall semester from September 2017 until January 2018 (five months). Although teachers were permitted to reseat students after randomization, we asked that students be moved in pairs to preserve deskmate assignments wherever possible. We defined the deskmate composition resulting from randomization and replacement as the intended seating chart for our intention-to-treat analysis.

Among recruited classrooms, we pre-registered to exclude all classrooms that did not meet certain inclusion criteria.¹⁴ We pre-registered a sample of 3,814 students. After additional pre-registered exclusions,¹⁵ our final analytic sample contained 3,365 students across 195 classrooms of 41 schools.

We measured twice the compliance with the intended seating chart. Teachers reported that 94 percent of students in the analytic sample sat next to their intended

deskmate in mid-September, two weeks after the start of the school year, and our field team recorded 86 percent compliance during in-person visits between October and December.

Data and Coding of Main Variables

The *primary treatment* variables derive from students' assignment to sit next to a specific, randomly allocated deskmate. We characterize the deskmate of student i in classroom c of school s using two baseline variables, T_{ics}^D : (1) deskmate's baseline GPA, defined as the average of deskmate's teacher-reported spring semester 2017 grades in the three core subjects (Hungarian literature, grammar, and mathematics, graded on a scale from 1 to 5, where 5 is best); and (2) deskmate's gender. *Secondary treatment variables* comprise deskmates' five subject-specific baseline grades in grammar, literature, mathematics, diligence, and behavior. We filled in missing teacher reports of students' baseline grades from students' self-reported baseline grades (3 percent of the cases).

Our *primary outcomes* are two measures of students' own educational achievement, Y_{ics} : (1) students' endline GPA (January 2018), defined as the average of students' teacher-reported end-of-term grades in the three core subjects Hungarian literature, grammar, and mathematics,¹⁶ (2) students' grade-specific standardized reading comprehension test score. Reading tests contained between 10 and 19 items, depending on grade level, and were developed for this study from the test banks of the PISA-like National Assessment of Basic Competencies (NABC) by the Hungarian Educational Authority. The test was administered in-class over 25-minutes as part of the 45-minute endline survey in the spring of 2018, one to three months after the completion of the intervention. Reading tests were machine-graded, precluding grader bias. We analyze the percentage of correctly answered items.

Secondary outcomes include (a) the five subject-specific fall-term grades in literature, grammar, mathematics, behavior, and diligence; and (b) the average of students' scores on teacher-written classroom tests in literature, grammar, and mathematics.¹⁷

We note that several of our outcomes depend on each other: endline GPA aggregates the subject-specific grades in literature, grammar, and mathematics; and these subject-specific grades, in turn, heavily depend on students' classroom-test scores and, to a smaller extent, on diligence and behavior. By contrast, our standardized reading test scores do not mechanically depend on other outcomes.

We collected additional *baseline information* on students' demographics (gender, age, ethnicity [Roma or not]) and socioeconomic status (asking teachers to name the richest/poorest students in their classrooms) from teacher reports. These covariates were used for balance checks. Anticipating missingness, we pre-registered to use these covariates only as robustness checks in supplementary outcomes models.

Table 1 gives descriptive statistics for the analytic sample. Half of the sample is female; students are 12 years old on average, and 25 percent are Roma. Baseline GPA had mean=3.7 and standard deviation=1. Missingness in baseline variables is low (0-5 percent). Moderate missingness in outcome reading scores (10 percent) is due to students' absence on the day of the endline survey or lack of parental

Table 1: Descriptive statistics for the main variables in the analytic sample.

	Mean	SD	Min	Max	N	% Missing
<i>Primary outcomes</i>						
Reading score (% correct)	54.65	20.85	0	100	3,022	10.19%
Endline GPA	3.57	1.03	1	5	3,241	3.68%
<i>Secondary outcomes (endline)</i>						
Grammar Grade	3.56	1.09	1	5	3,225	4.16%
Literature Grade	3.72	1.09	1	5	3,186	5.32%
Mathematics Grade	3.46	1.16	1	5	3,209	4.64%
Behavior Grade	4.19	0.90	2	5	3,243	3.63%
Diligence Grade	3.87	0.99	2	5	3,246	3.54%
Scores on teacher-written tests(%)	67.34	21.03	1	100	2,811	16.46%
<i>Baseline variables^a</i>						
Baseline GPA	3.69	1.01	1	5	3,272	2.76%
Grammar Grade	3.77	1.07	1	5	3,259	3.15%
Literature Grade	3.65	1.13	1	5	3,261	3.09%
Mathematics Grade	3.64	1.09	1	5	3,257	3.21%
Behavior Grade	4.28	0.84	2	5	3,273	2.73%
Diligence Grade	3.98	0.96	2	5	3,270	2.82%
Female	0.49	0.50	0	1	3,365	0.00%
Roma	0.25	0.43	0	1	3,192	5.14%
Poor	0.11	0.31	0	1	3,192	5.14%
Rich	0.10	0.30	0	1	3,192	5.14%
Age	11.91	1.84	7.82	17.49	3,314	1.52%
Grade level	5.33	1.67	3	8	3,365	0.00%

N=3,365. Grades and GPAs range from 1 to 5, where 5 is best.

^aMissingness in baseline grades and GPA is reported after replacing missing teacher reports with student self-reports, where possible.

consent to participate in the assessment. Missingness in fall-term test scores (16 percent) results from teacher non-response at endline.

Balance Checks

The key advantage of randomization is that it guarantees comparable (“balanced”) treatment and control groups in expectation and thus justifies causal inference without the threat of unobserved confounding. We tested for balance following Guryan, Kroft, and Notowidigdo (2009) by regressing each baseline variable, X_{ics} , of student i on students’ deskmates’ baseline characteristic, X_{ics}^D , the leave-one-out mean characteristic in the classroom, \bar{X}_{-ics} , and classroom fixed effects for the analytic samples of the two primary outcomes, with standard errors clustered at the school level. This procedure circumvents the artifactual correlation between student’s own and their deskmate’s characteristics induced by (successful) random partitioning within the classroom (Angrist 2014). Results demonstrate that the data are well-balanced (Online supplement Table A1). There are no, or only substantively minor, associations between students’ and their assigned deskmates’ baseline characteristics, and only one out of eleven associations is statistically significant.

Statistical Analyses

Primary Analyses

Our primary confirmatory analyses tested two main hypotheses. First, we hypothesized that sitting next to a deskmate with higher baseline achievement increases students' endline achievement (GPA or reading scores), Y_{ics} . Second, we hypothesized that sitting next to a girl increased students' endline achievement, Y_{ics} .

We evaluated each hypothesis using the following model,

$$Y_{ics} = \beta_0 + \beta_1 T_{ics}^D + \beta_2 T_{ics} + \delta X_{ics} + \eta_{cs} + \epsilon_{ics} \quad (1)$$

where T_{ics}^D stands either for deskmate's baseline GPA or deskmate's female gender, respectively; T_{ics} is student's own corresponding baseline characteristic, which must be controlled to avoid confounding by the artifactual correlation that is induced between students' and deskmates' characteristics by (successful) random partitioning (Angrist 2014); X_{ics} is a covariate to increase statistical precision (students' baseline GPA when T_{ics}^D is deskmate's gender, and student's gender when T_{ics}^D is deskmate's baseline GPA); η_{cs} are classroom fixed effects to account for the experimental design that randomized deskmates within classrooms; and ϵ_{ics} is an individual level error term.

Due to randomization, the focal coefficient β_1 identifies the causal effect of sitting next to a deskmate with characteristic T_{ics}^D .¹⁸ All other coefficients estimate nuisance parameters and do not have a causal interpretation.

Secondary Analyses

We executed five sets of exploratory secondary analyses. First, we replaced our primary outcomes in equation 1 with students' subject-specific endline grades and average scores from teacher-written classroom tests. Second, we replaced the outcome with students' subject-specific endline grades and the treatment with deskmates' corresponding subject-specific baseline variable.

Third, we explored heterogeneity in the effect of deskmate's baseline GPA by students' own baseline GPA. Specifically, following prior scholarship (for example, Duflo et al., 2011; Sacerdote, 2001), we divided baseline GPAs within each classroom into three tiers—low (bottom quartile), middle (middle two quartiles), and high (top quartile)—and fully interacted students' and deskmates' baseline GPA categories with each other,

$$Y_{ics} = a + \gamma_1 lL + \gamma_2 lM + \gamma_3 lH + \gamma_4 mL + \gamma_5 mH + \gamma_6 hL + \gamma_7 hM + \gamma_8 hH + \delta X_{ics} + \eta_{cs} + \epsilon_{ics} \quad (2)$$

where lower-case l, m, and h refer to student's own baseline GPA tiers, and upper-case L, M, and H, refer to their deskmate's baseline GPA tiers, with middle students sitting next to middle deskmates, mM , serving as the reference category.

Fourth, we estimated deskmate effects by deskmate's and student's own gender,

$$Y_{ics} = a + \varphi_1 bB + \varphi_2 gG + \varphi_3 gB + \delta X_{ics} + \eta_{cs} + \epsilon_{ics} \quad (3)$$

where g and b refer to the student's own gender (girl or boy), and G and B indicate deskmate's gender.¹⁹

Fifth, we explored the robustness of our results by replacing the preregistered measures of baseline and endline achievement with student scores on the comprehensive and nationally standardized NABC testing program. Since NABC scores became available only after preregistration, and only for select grade levels, we report results in Online supplement B.

Estimation, Standard Errors, and Multiple Hypothesis Testing

All models were estimated using the *xtreg* fixed-effects regression command in Stata, version 17 (StataCorp 2021). Robust standard errors are clustered at the school level. We report two types of hypothesis tests. First, we report two-sided hypothesis tests at conventional (unpenalized) levels of statistical significance. Second, we correct for multiple hypothesis testing and report penalized tests that hold the false-discovery rate at 5 percent using Benjamini and Hochberg's (1995) procedure. The false-discovery rate is the probability of making at least one "false-discovery" (Type I error) across a set of statistical tests. Firm norms for defining test sets have not yet evolved. We consider the deskmate-related tests within each model for a primary outcome, and again the deskmate-related tests across all models for secondary outcomes within each table or figure as a separate set of tests.²⁰

Results

Deskmate Effects by Deskmates' Baseline GPA

We found only weak evidence that sitting next to a deskmate with a higher (rather than a lower) baseline GPA increased the educational achievement of students on average. Although estimated deskmate effects were positive and in the expected direction for all primary and secondary outcomes (Table 2, and Online supplement Table A2), effect sizes were substantively small, and almost no estimate remained statistically significant after correcting for multiple hypothesis testing.

Specifically, the estimated causal effects of sitting next to a deskmate with a one-grade higher baseline GPA on the *primary outcomes*—student's own endline reading score ($\beta = 0.29$, $p = 0.45$) and endline GPA ($\beta = 0.02$, $p = 0.13$)—were only 1 percent and 2 percent of a standard deviation, respectively. Neither effect was statistically significant at the conventional 0.05 level (Table 2, Columns 1 and 2).

Columns 3 to 8 in Table 2 report exploratory results for the effects of sitting next to a deskmate with a higher baseline GPA on students' *secondary outcomes*. We found only weak evidence of deskmate-GPA effects on students' endline grades in literature, math, or behavior. Effects on grammar, diligence, and scores on teacher-written classroom tests were statistically significant at the conventional (unpenalized) 5-percent level but substantively small (not exceeding 4 percent of a standard deviation on any outcome).²¹ Only the estimated deskmate-GPA effect on students' endline grade in diligence remained statistically significant after correcting for multiple hypothesis testing, and even that only after adding pre-

Table 2: Deskmate Effects on Students' Achievement by Deskmate's Baseline GPA.

	Primary outcomes			Secondary outcomes				
	(1) Reading score	(2) Endline GPA	(3) Endline Grammar	(4) Endline Literature	(5) Endline Mathematics	(6) Endline Behavior	(7) Endline Diligence	(8) Scores on teacher-written tests
Own baseline GPA	9.14 [†] (0.50)	0.88 [†] (0.02)	0.86 [†] (0.02)	0.84 [†] (0.02)	0.94 [†] (0.02)	0.44 [†] (0.02)	0.74 [†] (0.02)	15.96 [†] (0.50)
Deskmate's baseline GPA	0.29	0.02	0.03*	0.00	0.03	0.03	0.04 [†]	0.71*
Female deskmate	(0.37)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.28)
Classroom fixed effects	2.87 [†] Yes	0.05 [†] Yes	0.18 [†] Yes	0.14 [†] Yes	-0.17 [†] Yes	0.43 [†] Yes	0.15 [†] Yes	-0.29 Yes
Mean of the outcome	54.91	3.58	3.56	3.72	3.47	4.20	3.88	67.31
Standardized deskmate effect	0.01	0.02	0.03	0.00	0.03	0.03	0.04	0.03
Observations	2,909	3,108	3,096	3,068	3,076	3,110	3,113	2,686
R-squared	0.23	0.74	0.63	0.61	0.62	0.33	0.58	0.70

[†] $p < 0.01$; * $p < 0.05$.

Note: GPA is computed from grades in grammar, literature, and mathematics (5 is best, 1 is worst). All models control for classroom fixed effects and include a constant. No additional controls are included. Cohen's D equals the deskmate coefficient divided by the standard deviation of the outcome. Robust standard errors are clustered at the school level in parentheses. Table reports conventional (unpenalized) hypothesis tests. Penalizing significance levels for multiple hypothesis testing by Benjamini and Hochberg's (1995) procedure with a 5% false-discovery rate across deskmate coefficients for the secondary outcomes indicates no statistically significant coefficients. Pre-registered robustness checks of each model with additional covariate controls in Appendix Table A2.

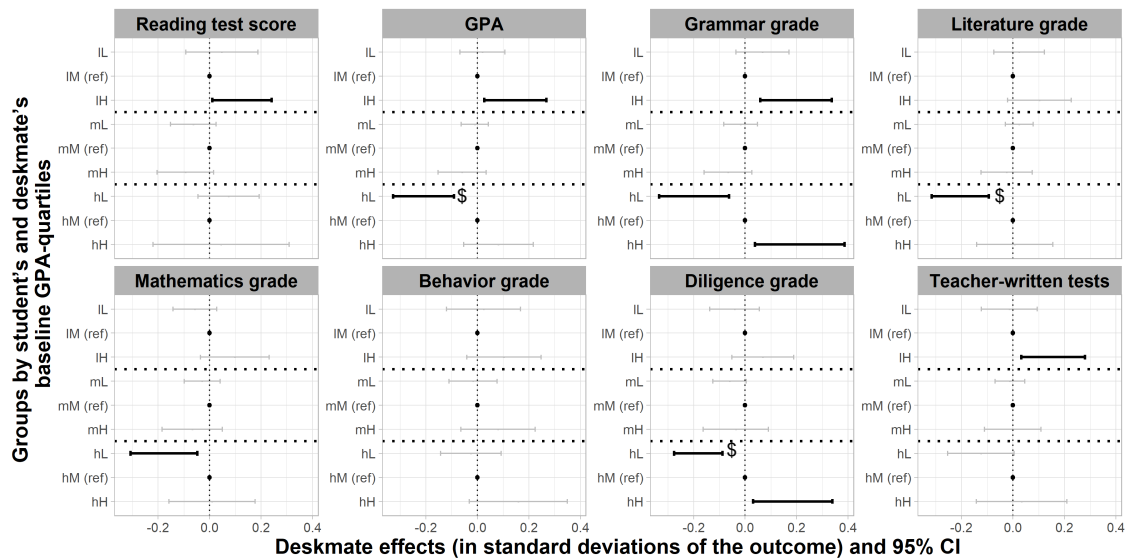


Figure 1: Heterogenous deskmate effects on educational outcomes by student's and deskmate's baseline GPAs. *Notes:* Estimated deskmate effects on various educational outcomes, by students' and deskmate's baseline achievement. Pre-registered specifications (Eq. 2). Each panel visualizes the effects of sitting next to a deskmate in the lowest (L) or highest (H) baseline-GPA quartile, respectively, compared to sitting next to a deskmate in the middle (M, reference) two quartiles, separately for students in the lowest (l), middle (m), and highest (h) baseline GPA quartile. Deskmate effects are expressed in standard deviations of the outcome. Bars display 95-percent confidence intervals. Bold bars indicate statistical significance at the (unpenalized) $\alpha=0.05$ level. For example, the faint bar in the top left corner shows that the effect of sitting next to a deskmate in the lowest baseline GPA quartile (compared to sitting next to a deskmate in the middle quartiles) on the endline reading score of students in the lowest baseline GPA quartile is not statistically significant. Contrasts marked by \$ remain statistically significant at a false-discovery rate of 5% after Benjamini-Hochberg (1995) corrections for multiple hypothesis testing (separately considering the tests within each model for a primary outcome [reading and GPA] and all tests across all models for secondary outcomes in this figure as a set, respectively). Full regression results shown in Appendix Table A5. Reference category: Desk-mate's baseline GPA = Middle

registered controls for additional baseline covariates (Online supplement Table A2, Column 7).

We conclude that the experiment does not provide dependable evidence for causal effects of sitting next to higher-achieving deskmates on the educational achievement of the average student. This conclusion is further supported by supplementary analyses of newly available comprehensive standardized test scores from the Hungarian NABC, a PISA-like testing program, as alternative exposure and outcome measures (see Online supplement B).

Because the average effects reported in Table 2 (and Tables A2 and A4) may obscure important effect heterogeneity, we next conducted heterogeneity analyses

that allowed deskmate effects to vary by deskmates' and students' own baseline GPA (Figure 1, equation 2). We found that sitting next to a high-GPA deskmate (rather than a middle-GPA deskmate) increased several achievement outcomes of low-GPA students (that is, reading scores, endline GPA, grammar grades, and scores on teacher-written tests). Conversely, sitting next to a low-GPA deskmate (rather than a middle-GPA deskmate) reduced high-GPA students' outcomes (that is, endline GPA, grammar, literature, mathematics and diligence grades, and scores on teacher-written tests). High-GPA students also benefited from sitting next to a high-GPA deskmate in grammar and diligence. Neither sitting next to a high-GPA nor a low-GPA deskmate (rather than a middle-GPA deskmate) appreciably affected any outcomes of middle-GPA students.²²

The positive effects of sitting next to a high-GPA deskmate for low-GPA students were similar in size to the negative effects of sitting next to a low-GPA deskmate for high-GPA students. Several of these heterogeneous effects were substantively quite large. For example, sitting next to a high-GPA deskmate increased low-GPA students' endline GPA by about 0.15 standard deviations, whereas sitting next to a low-GPA deskmate decreased high-GPA students' endline GPA by about 0.2 standard deviations.

Many of the conventionally statistically significant negative effects of sitting next to a low-GPA deskmate on high-GPA students' outcomes remained statistically significant after correcting for multiple hypothesis testing (that is, endline GPA, literature, diligence), as indicated in Figure 1, and several positive effects of sitting next to a high-GPA deskmate on low-GPA students' outcomes (that is, endline GPA, grammar, teacher-written tests) very nearly remained statistically significant, too (Online supplement Table A5).

We conclude that the experiment provides dependable evidence for strong heterogeneous deskmate effects on the educational achievement of high- and low-GPA students.

Deskmate Effects by Deskmates' Gender

We found no dependable evidence that sitting next to a female deskmate increased the educational achievement of the average student. Our *primary analyses* (Columns 1 and 2 in Table 3) show that sitting next to a girl (compared to a boy) has no detectable effect on the endline reading scores ($\beta = 0.03$, $p = 0.63$) or endline GPA ($\beta = 0.03$, $p = 0.08$) of students on average, as point estimates are positive but substantively small and not statistically significant. Our *secondary analyses* (Columns 3-8 in Table 3) indicate that sitting next to a girl increased only the average students' endline grammar grade ($\beta = 0.06$; $p = 0.02$) but not their other outcomes. All point estimates were substantively small and not statistically significant after correcting for multiple hypothesis testing.

Similarly, we found no dependable evidence for effect heterogeneity in the causal effect of sitting next to a girl either. The heterogeneity analysis shown in Figure 2 (equation 3) indicates positive effects of sitting next to a female deskmate on female students' endline GPA, grammar, and literature grades of around 5 percent of a standard deviation. These effects, however, were no longer statistically significant

Table 3: Deskmate Effects on Students' Achievement by Deskmate's Gender.

	Primary outcomes			Secondary outcomes				
	(1) Reading score	(2) Endline GPA	(3) Endline Grammar	(4) Endline Literature	(5) Endline Mathematics	(6) Endline Behavior	(7) Endline Diligence	(8) Scores on teacher-written tests
Female student	2.93 [†] (0.68)	0.06 [†] (0.02)	0.19 [†] (0.02)	0.15 [†] (0.02)	-0.16 [†] (0.03)	0.44 [†] (0.03)	0.15 [†] (0.02)	-0.24 (0.45)
Female deskmate	0.29 (0.57)	0.03 (0.02)	0.06* (0.02)	0.04 (0.02)	0.02 (0.02)	0.00 (0.03)	0.02 (0.02)	0.62* (0.30)
Own baseline GPA	9.07 [†] (0.52)	0.88 [†] (0.02)	0.86 [†] (0.02)	0.84 [†] (0.02)	0.94 [†] (0.02)	0.44 [†] (0.02)	0.73 [†] (0.02)	15.93 [†] (0.51)
Classroom fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of the outcome for students with male deskmate	54.20	3.54	3.52	3.68	3.43	4.19	3.85	66.59
Standardized deskmate effect	0.01	0.03	0.05	0.04	0.02	0.00	0.02	0.03
Observations	2,978	3,182	3,169	3,136	3,150	3,185	3,188	2,756
R-squared	0.23	0.74	0.63	0.61	0.62	0.33	0.58	0.71

[†] $p < 0.01$; * $p < 0.05$.

Note: GPA is computed from grades in grammar, literature, and mathematics (5 is best, 1 is worst). All models control for classroom fixed effects and include a constant. No additional controls are included. Cohen's D equals the deskmate coefficient divided by the standard deviation of the outcome. Robust standard errors are clustered at the school level in parentheses. Table reports conventional (unpenalized) hypothesis tests. Penalizing significance levels for multiple hypothesis testing by Benjamini and Hochberg's (1995) procedure with a 5% false-discovery rate across deskmate coefficients for the secondary outcomes indicates no statistically significant coefficients. Corresponding models with additional covariate controls in Appendix Table A3.

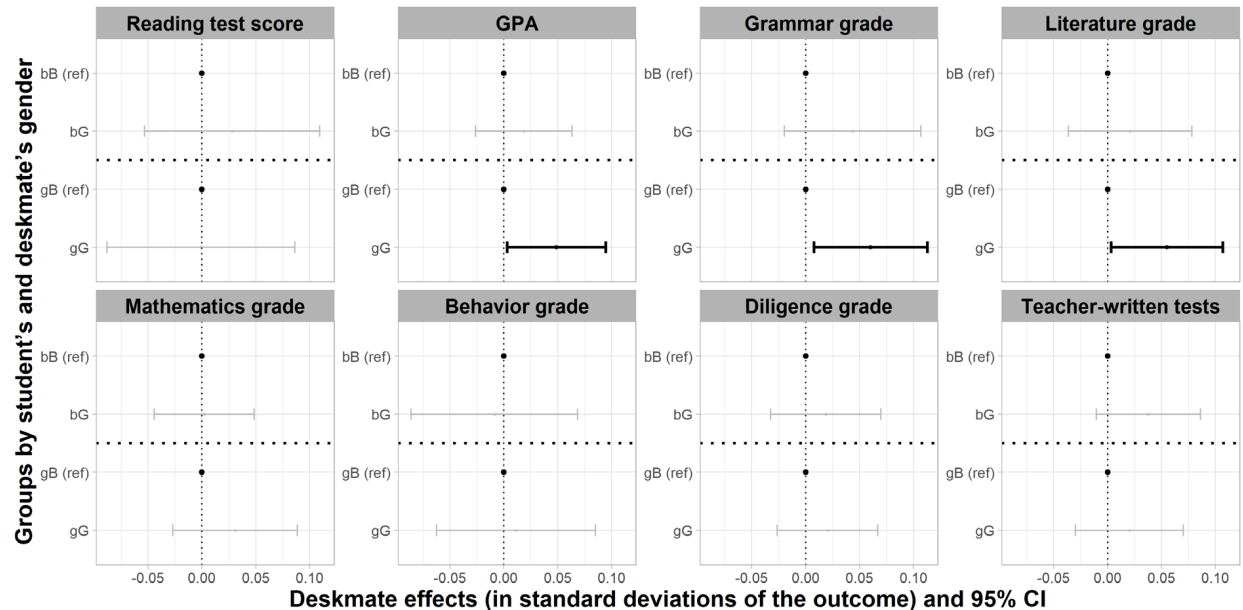


Figure 2: Heterogenous deskmate effects on educational outcomes by student's and deskmate's gender.

Notes: Estimated deskmate effects on various educational outcomes, by students' and deskmate's gender. Pre-registered specifications (Eq. 3). Each panel visualizes the effects of sitting next to female (G) rather than a male (B) deskmate, separately for male (b) and female (g) students. Deskmate effects are expressed in standard deviations of the outcome. Bars display 95-percent confidence intervals. Bold bars indicate statistical significance at the (unpenalized) $\alpha=0.05$ level. No estimate remains statistically significant after Benjamini-Hochberg (1995) corrections for multiple hypothesis testing at a false-discovery rate of 5% (separately considering the tests within each model for a primary outcome [reading and GPA] and all tests across all models for secondary outcomes in this figure as a set, respectively). Full regression results shown in Appendix Table A6. Reference category: Desk-mate's gender = Male

after correcting for multiple hypothesis testing. There is no evidence that boys benefit from (or are harmed by) sitting next to girls.

Consequences for aggregate achievement levels and inequality

The presence of heterogenous deskmate effects across students with different baseline GPAs (Figure 1) implies that changing the seating chart can affect aggregate achievement levels and inequality. We illustrate the effects on within-classroom achievement levels and inequality with stylized simulations (assuming classrooms containing 24 students seated at freestanding two-person desks).²³ The key simulation parameters are the heterogenous causal deskmate effects estimated from equation 2, shown in Figure 1. We further assume that students differ only with respect to their baseline GPAs (mean=3.8, SD=0.9 on a 5-point scale). We consider the effects of three prototypical seating charts: achievement-concordant seating charts that pair-up students with the most similar baseline GPAs; achievement-discordant seating charts that pair-up high-GPA with low-GPA students; and random seating charts that allocate deskmates randomly.²⁴

Table 4: Simulated effects of different seating chart designs on levels and within-classroom inequality in student achievement (GPA).

	Mean endline GPA	Mean ratio relative to random seating	Standard deviation endline GPA	Standard-deviation ratio relative to random seating
PANEL A: ALL STUDENTS				
Random seating chart	3.75		0.69	
Achievement-concordant seating chart	3.78	1.01	0.73	1.06
Achievement-discordant seating chart	3.74	1.00	0.59	0.86
PANEL B: HIGH BASELINE GPA STUDENTS (TOP 25%)				
Random seating chart	4.67		0.12	
Achievement-concordant seating chart	4.79	1.03	0.05	0.50
Achievement-discordant seating chart	4.5	0.96	0.05	0.42
PANEL C: MIDDLE BASELINE GPA STUDENTS (MIDDLE 50%)				
Random seating chart	3.79		0.06	
Achievement-concordant seating chart	3.81	1.01	0.05	0.83
Achievement-discordant seating chart	3.81	1.01	0.05	0.83
PANEL D: LOW BASELINE GPA STUDENTS (BOTTOM 25%)				
Random seating chart	2.74		0.08	
Achievement-concordant seating chart	2.72	0.99	0.05	0.63
Achievement-discordant seating chart	2.85	1.04	0.05	0.63

Note: Table shows averages of 500 simulation runs for a classroom with 24 students seated at two-person desks and a baseline GPA distribution that mimics the empirical distribution (left-skewed, min=1, mean=3.8, max=5, SD = 0.9); simulation parameters for causal deskmate effects on endline GPA are from Eq. 2

Table 4 presents three main results.²⁵ First, rearranging the seating chart by prior achievement has no appreciable effect on the average achievement level within the classroom. In the random seating chart, the average endline GPA is 3.75 (SD=0.69). In the achievement-concordant and achievement-discordant seating charts, the average achievements are 3.78 and 3.74, respectively. These differences are too small to matter in practice.²⁶ Rearranging the seating chart has no appreciable effect on achievement levels because the substantial deskmate effects on the fifty percent of students with low or high prior achievement are still too small to compensate for the absence of deskmate effects on the middle half of students.

Second, rearranging the seating chart by prior achievement does affect aggregate inequality within classrooms. In the random seating chart, the standard deviation of students' endline GPAs is 0.69 grade points. Compared to random seating, achievement-concordant seating increases the within-classroom standard deviation

by 6 percent to 0.73. By contrast, achievement-discordant seating decreases the standard deviation by 14 percent to 0.59 (Panel A).

It follows from the parameter estimates shown in Figure 1 that all movement is in the tails of the distribution (Column 2, Panels B-D). Compared to random seating, achievement-discordant seating reduces the achievement of students with high baseline GPA by -0.17 points, because they now exclusively sit next to low-achieving deskmates, which exert a negative effect on their grades. The same seating chart increases the achievement of low-GPA students by +0.1 points, because they now exclusively sit next to high-GPA deskmates, which exert a positive effect on their grades. The reverse explains the inequality-increasing effect of achievement-concordant seating: no high-GPA students sit next to low-GPA students, which would have decreased their achievement, and no low-GPA students sit next to high-GPA students, which would have increased their achievement.

Third, as an aside, we note that seating charts also affect inequality *within* each tier of students: compared to random seating, both achievement-concordant and achievement-discordant seating charts decrease the variation of endline outcomes *within* the groups of students with high-, middle-, and low baseline GPA (Panels B-D, Columns 3 and 4). The reason is that, compared to random seating, the two systematic seating charts expose each group of students to a more homogenous group of deskmates.²⁷ This demonstrates that systematic seating charts based on students' baseline characteristics can increase the internal coherence of groups defined by these characteristics, regardless of whether the seating chart increases or decreases outcomes inequality within the classroom overall.

Discussion

Peer effects matter because they can influence educational choices and outcomes. Dependable, field-experimental evidence for peer effects, however, remains scarce, especially at the sub-classroom level. We conducted a large field experiment to investigate causal peer effects of deskmates on students' learning outcomes by randomizing the seating chart in 195 classrooms for the duration of one semester. We draw two main conclusions. First, there is no dependable evidence for causal effects of sitting next to deskmates with high baseline GPAs or next to girls on students' average learning outcomes. Although estimates pointed in the expected positive direction across all outcomes (that is, standardized reading scores, endline GPA, subject-specific grades, and classroom exams), point estimates were substantively small, few estimates were statistically significant by conventional standards, and hardly any remained statistically significant after correcting for multiple hypothesis testing. Because our sample was large and standard errors were small, we can rule out substantively large deskmate effects on the average student's learning outcomes.

Our results thus contrast with prior observational studies that found large average deskmate effects in an elite Korean university (Hong and Lee 2017) and in Hungarian high schools (Keller and Takács 2019). Our results also fail to confirm Lu and Anderson's (2015) experimental finding of the positive effects of sitting next to a girl in one Chinese middle school. Like the randomized experiments in China

(Lu and Anderson 2015); Li et al. 2014, Wu et al. 2023), we find no evidence for an effect of sitting next to a higher-achieving deskmate on average..

Second, we found dependable evidence for substantial effect heterogeneity by students' own baseline achievement (GPA), similar to Hong and Lee's (2017) observational findings. Across multiple learning outcomes, the lowest-achieving students benefited from sitting next to the highest-achieving students in the classroom. In turn, multiple outcomes of the highest-achieving students were diminished by sitting next to the lowest-achieving students. For example, sitting next to a deskmate from the bottom quartile of students within the classroom (measured by baseline GPA) reduced the endline GPA of top quartile students by 0.2 standard deviations, which is nearly as much as the descriptive difference in baseline GPA between boys and girls. Concretely, this means that two out of three high-achieving students seated next to low-achieving deskmates rather than middle-achieving deskmates would receive a lower grade in one of the three core subjects that comprise their GPA in our analysis.²⁸

Heterogenous peer effects have important policy implications. Although we found that seating charts do not meaningfully affect aggregate achievement levels (efficiency), they can affect inequality between students: Achievement-concordant seating charts would increase inequality, whereas achievement-discordant seating would decrease inequality. In this respect, our experimental results align with the observational literature on between- and within-school tracking, which reports that tracking by prior achievement (as the macroscopic equivalent of achievement-concordant seating charts) is associated with increased inequality (Terrin and Triventi 2022), but contrast with experimental studies that sometimes indicate that tracking can increase aggregate achievement levels (for example, Duflo et al. 2011).

Employing seating charts to influence inequality presents ethical questions that hinge on competing policy objectives and normative preferences regarding which students should benefit. For example, policymakers who want to decrease inequality in student achievement might advocate achievement-discordant seating charts. The resulting decrease in inequality, assuming that our results generalize to the new setting, however, would be accomplished not only by increasing the achievement of the lowest-achieving students but also by reducing the achievement of the highest-achieving students. The gain of one group of students would be the other group's loss. Furthermore, the highest-achieving students within a given classroom will often be distinctly underprivileged when schools are stratified by prior achievement (for example, due to regional and neighborhood differences, as in the United States and Hungary). Reducing aggregate inequality by adopting achievement-discordant seating charts might thus unintentionally hurt the most promising among the neediest students. By contrast, policymakers who want to avoid systematically harming specific groups of students (or, because all students have to sit next to somebody, spread inevitable harm and benefits fairly) might advocate for frequent random reseating, albeit at the cost of foregoing the inequality-reducing effect of achievement-discordant seating.

Whether seating charts can effectively influence inequality also depends on additional parameters. Two considerations give pause. First, peer effects typically

are ‘second-order’ effects (Borgen, Borgen, and Birkelund 2023). In ours as well as all prior deskmate studies, their magnitudes pales in comparison to the educational inequalities indexed by students’ own and their families’ baseline characteristics. Second, changing the seating chart will affect total inequality within a school, district, or nation only indirectly via its effect on within-classroom inequality. Consequently, the relative effect of seating charts on total inequality will diminish with the magnitude of between-classroom inequality. If there is no between-classroom inequality in students’ baseline characteristics, then the effect of the seating chart on within-classroom inequality equals its effect on total inequality and hence will be large. If, however, students’ achievement differs across classrooms, for example, due to geographic differences or ability tracking, then the effect of the seating chart on total inequality will be smaller.

That said, decades of costly, controversial, and frequently failed policy interventions at supra-classroom levels have demonstrated that changing aggregate educational outcomes is exceedingly difficult. Although their effects may be small at the individual level, seating charts are a promising tool for affecting educational inequality in the aggregate because intervening on seating charts is eminently feasible in practice.

Notes

- 1 For example, Duflo et al.’s (2011) famous tracking experiment in Kenyan primary schools found that all students benefited from exposure to higher-achieving peers, which would suggest negative effects of tracking on low-achieving students. Nevertheless, the total effect of tracking was positive for all students because teachers adjusted their teaching style to benefit low-achieving students.
- 2 Because college roommates rarely take the same courses (exceptions include Carrell, Fullerton, and West 2009; Lyle 2007; Li et al. 2019), this literature speaks more to the effect of coresidence (akin to neighborhood effects) than to peer effects in the context of receiving shared instruction (Stinebrickner and Stinebrickner 2006).
- 3 Cheating affects students’ measured learning outcomes (grades and test scores). Regardless of whether students learn anything in the process, measured outcomes are important because they affect grade retention and a host of future transitions (e.g., educational enrollment, hiring). We imply no normative judgement.
- 4 Another potential mechanisms is raised by ‘Big-Fish-Little-Pond’ theory in educational psychology, which posits that exposure to academically stronger peers depresses, and exposure to academically weaker peers increases students’ academic self-concept, which, in turn, might affect students’ achievement (e.g., Marsh and Parker 1984). We tested this prediction in the current setting and found no evidence for deskmate effects on students’ academic self-concept (Keller, Kim, and Elwert 2023).
- 5 In observational studies of peer effects, classical measurement error produces upward bias rather than attenuation toward zero (Angrist, 2014). The attenuating effect of classical measurement error is restored when peer relations are randomly assigned (Feld and Zölitz 2017).
- 6 The authors argue against differential disruptiveness as a mediator. Copying from a deskmate on the outcomes test was prevented by spreading test takers across multiple rooms.

- 7 Within the literature on peer effects in education, only research on tracking extensively discusses macro-level implications for inequality (e.g., Card and Giuliano 2016; Hanushek and Wößmann 2006; Oakes 1985).
- 8 Whether Hong and Lee's (2017) finding of non-linearities in the effect of sitting next to a higher-achieving deskmate (larger in the tails than in the middle of the achievement distribution) have implications for aggregate achievement levels is uncertain, because the differences are small.
- 9 Based on reports from a 2017 survey of N=160 homeroom teachers in our experimental sample regarding their own prior seating chart practices.
- 10 In 2021, we conducted a national survey of N=413 primary-school homeroom teachers. 74 percent of homeroom teachers employed a fixed seating chart. The majority (53 percent) of these teachers indicated changing the seating chart at most once per school year; 10 percent changed the seating chart at least monthly.
- 11 See http://eduline.hu/kozoktatas/2018/8/27/mit_tanulnak_a_diakok_az_iskolaban_98J42H and http://eduline.hu/kozoktatas/2018/8/28/felsos_kerettanternv_2018_Q86D33.
- 12 In a 2020 survey of N=135 teachers, 3 out of 10 teachers reported giving students worse grades for "talking or laughing," "shouting," or "walking around" during lessons, and 6 out of 10 reported giving worse grades for "playing or reading something else" during lessons.
- 13 "Grading on a curve" describes the practice of fixing certain features of the grade distribution ("curve") in advance, perhaps to comply with stakeholder expectations, tradition, or to limit grader effects. Clearly, it is impossible to detect average effects on grades when the mean of the grade distribution is pre-determined, and it is impossible to detect effects on inequality when the entire grade distribution is pre-determined.
- 14 We dropped 120 classrooms for the following reasons: withdrawal from the study prior to the start of the fall semester (25); containing fewer than 10 students at baseline (8); split classrooms in any of the three focal subjects (10); classroom layouts not comprising free-standing two-person desks (30); unreliable baseline reporting (7); failure to implement the seating chart (40). We note that dropping classrooms does not affect internal validity because randomization occurred within classrooms so that each classroom constitutes a separate randomized experiment.
- 15 Pre-registration had counted some students twice, reducing its size to 3,803 unique students. Five more students were excluded because their classroom was smaller than the pre-registered minimum class size of 10; and 36 students from the spring rosters had left their classrooms before the start of the fall semester. As pre-registered, we further deleted 397 students who were randomized to sit alone.
- 16 We computed GPA from non-missing grades. We did not impute outcomes.
- 17 We also pre-registered the average of students' monthly fall-semester grades as a secondary outcome. Because this variable is essentially identical to the primary outcome (students' fall semester GPA, $\rho = 0.95$), we do not show results. The analytical scripts for Table 2 and Table 3 produce all results. Analyses for whether students liked a school subject are reported in Keller et al. (2023).
- 18 Note that the experiment randomizes deskmates, not deskmates' characteristics. Therefore, we estimate the causal effect of sitting next to deskmates who have certain characteristics, not the effect of intervening on the characteristics themselves (Bramoullé, Djebbari, and Fortin 2020; Ogburn and VanderWeele 2014).

- 19 We did not execute pre-registered heterogeneity analyses using machine learning tools, because the software packages we pre-registered cannot accommodate the classroom effects fixed that our analysis requires for identification.
- 20 In so doing, we attempt to strike a balance between over- and under-penalization. On one hand, because Benjamini and Hochberg's (1995) procedure does not account for dependencies between tests, the procedure is conservative (that is, reduces the power to reject the null of no effect). In this sense, our corrections may be considered stringent. On the other hand, combining test sets across tables and figures would make the corrections even more conservative. In this sense, our choice not to declare test sets across tables and figures may be considered lenient.
- 21 The deskmate effects on students' endline grades in grammar and diligence appear to be driven by deskmates' baseline grades in grammar and diligence, respectively (Online supplement Table A4)—although these estimates are not significant after correcting for multiple hypothesis testing, either.
- 22 This suggests that the absence of effects on the learning outcomes of the average student (reported in Table 2) is explained by the absence of effects on middle-GPA students, who comprise half of the sample.
- 23 We defer the formal analysis of optimal seating chart assignments to future research.
- 24 Achievement-concordant seating charts rank students by GPA and seat pairs of most-similar students at the same desk. Achievement-discordant seating charts rank students by baseline GPA and seat the student with the highest GPA next to the student with the lowest GPA, and then continue to form subsequent pairs in order. In each classroom, there exist one achievement-concordant, one achievement-discordant, and many possible random seating charts. Table 4 reports averages across 500 simulated classrooms. We also considered various other seating charts, which did not add much and are omitted here.
- 25 Simulation results should primarily be read for qualitative evidence on the direction of effects. The quantitative change in inequality depends on additional parameters, which this simulation does not vary. For example, the absolute effects of seating charts on within-classroom inequality will generally increase with the variance in the baseline-characteristic on which the seating chart is based. Pointedly, even with strongly heterogeneous deskmate effects, changing deskmates will not change outcomes when students do not differ in their baseline characteristics.
- 26 Neither achievement-discordant nor achievement-concordant seating charts guarantee the most extreme results in general.
- 27 Concretely, under random seating, high-GPA students are exposed to a mix of low, middle, and high baseline GPA deskmates, which pull their endline GPA in all directions. By contrast, in achievement-concordant (discordant) seating, the endline GPA of all students with high baseline GPA is pulled upward (downward), because all have high-(low-)GPA deskmates, resulting in less endline inequality within the high-achievement group. Similar reasoning explains the inequality-decreasing effect of systematic seating charts for students with low and middle baseline GPAs, respectively.
- 28 The aggregate endline GPA has $SD \approx 1$. Reducing the endline grade in one out of three subjects by 1 for two out of three students reduces endline GPA by $\frac{2}{3}/3 \approx 0.2$.

References

- Angrist, Joshua D. 2014. "The Perils of Peer Effects." *Labour Economics* 30:98–108. <https://doi.org/10.1016/j.labeco.2014.05.008>

- Antecol, Heather, Ozkan Eren, and Serkan Ozbeklik. 2016. "Peer Effects in Disadvantaged Primary Schools Evidence from a Randomized Experiment." *Journal of Human Resources* 51(1):95–132. <https://doi.org/10.3368/jhr.51.1.95>
- Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing Author." *Journal of the Royal Statistical Society. Series B* 57(1):289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Booij, Adam S., Edwin Leuven, and Hessel Oosterbeek. 2016. "Ability Peer Effects in University: Evidence from a Randomized Experiment." *The Review of Economic Studies* rdw045. <https://doi.org/10.1093/restud/rdw045>
- Borgen, Nicolai T., Solveig T. Borgen, and Gunn Elisabeth Birkelund. 2023. "The Counteracting Nature of Contextual Influences: Peer Effects and Offsetting Mechanisms in Schools." *Social Forces* 101(3):1288–1320. <https://doi.org/10.1093/sf/soac023>
- Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. 2020. "Peer Effects in Networks: A Survey." *Annual Review of Economics* 12:603–29. <https://doi.org/10.1146/annurev-economics-020320-033926>
- Bygren, Magnus. 2016. "Ability Grouping's Effects on Grades and the Attainment of Higher Education." *Sociology of Education* 89(2):118–36. <https://doi.org/10.1177/0038040716642498>
- Card, David, and Laura Giuliano. 2016. "Can Tracking Raise the Test Scores of High-Ability Minority Students?" *American Economic Review* 106(10):2783–2816. <https://doi.org/10.1257/aer.20150484>
- Carrell, Scott E., Richard L. Fullerton, and James E. West. 2009. "Does Your Cohort Matter? Measuring Peer Effects in College Achievement." *Journal of Labor Economics* 27(3):439–64. <https://doi.org/10.1086/600143>
- Coleman, James S. 1966. *Equality of Educational Opportunity Report*. Washington, DC: US Government Printing Office.
- Coleman, James S. 1968. "Equality of Educational Opportunity." *Equity & Excellence in Education* 6(5):19–28. <https://doi.org/10.1080/0020486680060504>
- Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review* 101(August):1739–74. <https://doi.org/10.1257/aer.101.5.1739>
- Elwert, Felix, and Christopher Winship. 2014. "Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable." *Annual Review of Sociology* 40:31–53. <https://doi.org/10.1146/annurev-soc-071913-043455>
- Feld, Jan, and Ulf Zölitz. 2017. "Understanding Peer Effects: On the Nature, Estimation, and Channels of Peer Effects." *Journal of Labor Economics* 35(9448):000–000. <https://doi.org/10.1086/689472>
- Foster, Gigi. 2006. "It's Not Your Peers, and It's Not Your Friends: Some Progress toward Understanding the Educational Peer Effect Mechanism." *Journal of Public Economics* 90(8–9):1455–75. <https://doi.org/10.1016/j.jpubeco.2005.12.001>
- Guryan, Jonathan, Kory Kroft, and Matthew J. Notowidigdo. 2009. "Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments." *American Economic Journal: Applied Economics* 1(4):34–68. <https://doi.org/10.1257/app.1.4.34>
- Hallinan, Maureen T. 1994. "Tracking: From Theory to Practice." *Sociology of Education* 67(2):79.
- Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rivkin. 2003. "Does Peer Ability Affect Student Achievement?" *Journal of Applied Econometrics* 18(5):527–44. <https://doi.org/10.1002/jae.741>

- Hanushek, Eric A., and Ludger Wößmann. 2006. "Does Educational Tracking Affect Performance and Inequality? Differences- In-Differences Evidence Across Countries." *The Economic Journal* 116(510):C63–76. <https://doi.org/10.1111/j.1468-0297.2006.01076.x>
- Hong, Sok Chul, and Jungmin Lee. 2017. "Who Is Sitting next to You? Peer Effects inside the Classroom." *Quantitative Economics* 8(1):239–75. <https://doi.org/10.3982/qe434>
- Keller, Tamás, Jinho Kim, and Felix Elwert. 2023. "Peer Effects on Academic Self-Concept: A Large Randomized Field Experiment." *European Sociological Review* (January):1–16. <https://doi.org/10.1093/esr/jcad001>
- Keller, Tamás, and Károly Takács. 2019. "Peers That Count: The Influence of Deskmates on Test Scores." *Research in Social Stratification and Mobility* 62(May):100408. <https://doi.org/10.1016/j.rssm.2019.05.003>
- Lazear, E. P. 2001. "Educational Production." *The Quarterly Journal of Economics* 116(3):777–803.
- Li, Tao, Li Han, Linxiu Zhang, and Scott Rozelle. 2014. "Encouraging Classroom Peer Interactions: Evidence from Chinese Migrant Schools." *Journal of Public Economics* 111: 29–45. <https://doi.org/10.1016/j.jpubeco.2013.12.014>
- Li, Xinran, Peng Ding, Qian Lin, Dawei Yang, and Jun S. Liu. 2019. "Randomization Inference for Peer Effects." *Journal of the American Statistical Association* 114(528):1651–64. <https://doi.org/10.1080/01621459.2018.1512863>
- Lu, Fangwen, and Michael L. Anderson. 2015. "Peer Effects in Microenvironments: The Benefits of Homogeneous Classroom Groups." *Journal of Labor Economics* 33(1):91–122. <https://doi.org/10.1086/677392>
- Lyle, David S. 2007. "Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point." *Review of Economics and Statistics* 89(2):289–99. <https://doi.org/10.1162/rest.89.2.289>
- Marsh, Herbert W., and John W. Parker. 1984. "Determinants of Student Self-Concept: Is It Better to Be a Relatively Large Fish in a Small Pond Even If You Don't Learn to Swim as Well?" *Journal of Personality and Social Psychology* 47(1):213–31. <https://doi.org/10.1037/0022-3514.47.1.213>
- McEwan, Patrick J., and Kristen A. Soderberg. 2006. "Roommate Effects on Grades: Evidence from First-Year Housing Assignments." *Research in Higher Education* 47(3):347–70. <https://doi.org/10.1007/s11162-005-9392-2>
- Oakes, Jeannie. 1985. *Keeping Track: How Schools Structure Inequality*. New Haven: Yale University Press.
- Ogburn, Elizabeth L., and Tyler J. VanderWeele. 2014. "Causal Diagrams for Interference." *Statistical Science* 29(4):559–78. <https://doi.org/10.1214/14-sts501>
- Oosterbeek, Hessel, Nienke Ruijs, and Inge de Wolf. 2023. "Heterogeneous Effects of Comprehensive vs. Single-Track Academic Schools: Evidence from Admission Lotteries." *Economics of Education Review* 93(January):102363. <https://doi.org/10.1016/j.econedurev.2023.102363>
- Rohrer, Julia M., Tamás Keller, and Felix Elwert. 2021. "Proximity Can Induce Diverse Friendships: A Large Randomized Classroom Experiment" edited by F. M. Origo. *PLOS ONE* 16(8):e0255097. <https://doi.org/10.1371/journal.pone.0255097>
- Sacerdote, Bruce I. 2001. "Peer Effects with Random Assignment: Results for Dartmouth Roommates." *The Quarterly Journal of Economics* 116(2):681–704. <https://doi.org/10.1162/00335530151144131>

- Schiltz, Fritz, Deni Mazrekaj, Daniel Horn, and Kristof De Witte. 2019. "Does It Matter When Your Smartest Peers Leave Your Class? Evidence from Hungary." *Labour Economics* 59(October 2018):79–91. <https://doi.org/10.1016/j.labeco.2019.04.001>
- Shalizi, Cosma Rohilla, and Andrew C. Thomas. 2011. "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies." *Sociological Methods & Research* 40(2):211–39. <https://doi.org/10.1177/0049124111404820>
- Slavin, Robert E. 1990. "Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis." *Review of Educational Research* 60(3):471–99. <https://doi.org/10.3102/00346543060003471>
- StataCorp. 2021. "Stata Statistical Software."
- Stinebrickner, Ralph, and Todd R. Stinebrickner. 2006. "What Can Be Learned about Peer Effects Using College Roommates? Evidence from New Survey Data and Students from Disadvantaged Backgrounds." *Journal of Public Economics* 90(8–9):1435–54. <https://doi.org/10.1016/j.jpubeco.2006.03.002>
- Terrin, Éder, and Moris Triventi. 2023. "The Effect of School Tracking on Student Achievement and Inequality: A Meta-Analysis." *Review of Educational Research* 93(2):236–74. <https://doi.org/10.3102/00346543221100850>
- Wu, Jia, Junsen Zhang, and Chunchao Wang. 2023. "Student Performance, Peer Effects, and Friend Networks: Evidence from a Randomized Peer Intervention." *American Economic Journal: Economic Policy* 15(1): 510–542. <https://doi.org/10.1257/pol.20200563>

Acknowledgments: The authors thank Carlo Barone, Dorottya Baross, Steven Durlauf, Edina Gábor, Eric Grodsky, Judit Kerek, Gábor Kertesi, Gábor Kézdi, Andreas Kot-sadam, Xinran Li, Károly Takács, Jeffrey Smith, and audiences at the University of Wisconsin-Madison, the University of Chicago, ISA RC28 Spring Meeting in Frankfurt am Main, the Annual Meeting of the International Network of Analytical Sociologists (INAS) in St. Petersburg, the Meeting of the Economics of Education Association (AEDE) in Las Palmas, the Università Cattolica del Sacro Cuore, Milan, and the European Research Network on Transitions in Youth (TiY) in Mannheim for valuable discussions. This research was funded by grants from the Hungarian National Research, Development, and Innovation Office (NKFIH), grant number K-135766; a János Bolyai Research Scholarship of the Hungarian Academy of Sciences, #BO/00569/21/9; the New National Excellence Program of the Ministry for Culture and Innovation, # ÚNKP-23-5-CORVINUS-149; and a Romnes Fellowship and a Vilas Midcareer Faculty Award, University of Wisconsin-Madison. Direct correspondence to Tamás Keller (keller.tamas@tk.hu) and Felix Elwert (elwert@wisc.edu).

Tamás Keller: HUN-REN Centre for Social Sciences, Computational Social Science - Research Center for Educational and Network Studies, Budapest, Hungary
HUN-REN Centre for Economic and Regional Studies, Institute of Economics, Budapest, Hungary.
TÁRKI Social Research Institute, Budapest, Hungary. E-mail: keller.tamas@tk.hu.

Felix Elwert: Department of Sociology & Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison. E-mail: elwert@wisc.edu.