INSTITUTE OF ECONOMICS
Centre for Economic and Regional Studies

Közgazdaság- és Regionális Tudományi Kutatóközpont
KÖZGAZDASÁG-TUDOMÁNYI INTÉZET

KRTK-KTI WORKING PAPERS | KRTK-KTI MŰHELYTANULMÁNYOK

Peer Effects on Academic Self-concept: A Large Randomized Field Experiment

TAMÁS KELLER – JINHO KIM – FELIX ELWERT

KRTK-KTI WP – 2022/25

December 2022

https://kti.krtk.hu/wp-content/uploads/2022/12/CERSIEWP202225.pdf

INSTITUTE OF ECONOMICS, CENTRE FOR ECONOMIC AND REGIONAL STUDIES
EÖTVÖS LORÁND RESEARCH NETWORK (ELKH)
BUDAPEST, 2022

ABSTRACT

Social theories posit that peers affect students' academic self-concept (ASC). Most prominently, Big-Fish-Little-Pond, invidious comparison, and relative deprivation theories predict that exposure to academically stronger peers decreases students' ASC, and exposure to academically weaker peers increases students' ASC. These propositions have not yet been tested experimentally. We executed a large and pre-registered field experiment that randomized students to deskmates within 195 classrooms of 41 schools (N = 3,022). Our primary experimental analysis found no evidence of an effect of peer achievement on ASC in either direction. Exploratory analyses hinted at a subject-specific deskmate effect on ASC in verbal skills, and that sitting next to a lower-achieving boy increased girls' ASC (but not that sitting next to a higher-achieving boy decreased girls' ASC). Critically, however, none of these group-specific results held up to even modest corrections for multiple hypothesis testing. Contrary to theory, our randomized field experiment thus provides no evidence for an effect of peer achievement on students' ASC.

JEL codes: C93, I21, I24

Keywords: Academic self-concept, peer effects, social comparison, Big-Fish-Little-Pond, invidious comparison, relative deprivation, randomized field experiment, deskmates, Hungary

Tamás Keller
KRTK KTI; Computational Social Science - Research Center for Educational and Network Studies, Centre for Social Sciences; and TÁRKI Social Research Institute
e-mail: keller.tamas@tk.hu

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

Jinho Kim
Department of Health Policy and Management, Korea University and Interdisciplinary Program in Precision Public Health, Korea University

Acknowledgement:
Tamás Keller and Felix Elwert contributed equally to this research.

Funding:
The research was funded by grants from the Hungarian National Research, Development, and Innovation Office (NKFIH), FK-125358; a János Bolyai Research Scholarship, Hungarian Academy of Sciences, BO/00569/21/9; the ÚNKP-22-5-CORVINUS-134 New National Excellence Program, Ministry for Innovation and Technology; and a Romnes Fellowship, University of Wisconsin-Madison.

A kortársak hatása az akadémiai önképre: Egy nagymintás randomizált terepkísérlet

KELLER TAMÁS – JINHO KIM – FELIX ELWERT

ÖSSZEFOGLALÓ

Többféle társadalmi elmélet szerint is a kortársak hatással vannak a diákok akadémiai önértékelésére, vagyis arra, hogy valaki mennyire gondolja jónak magát egy adott tantárgyból. A Big-Fish-Little-Pond, a hátrányos összehasonlítás és a relatív depriváció elméletek alapján azt feltételezhetjük, hogy a jobban teljesítő osztálytársak csökkentik, a rosszabbul teljesítők pedig növelik a diákok akadémiai önértékelését. Ugyanakkor erre az elméleti felvetésre egyelőre nincsenek randomizált kísérletekből származó bizonyítékaink. Tanulmányunk ezt a hiányt tölti be. Egy nagymintás, előregisztrált terepkísérletet végeztünk, amely során 41 iskola 195 osztályában (N = 3 022) véletlenszerű ültetési rendet alkalmaztunk, vagyis a diákok véletlenül kisorsolt padtársak mellett ültek. Eredményeink szerint a padtársak semmilyen irányban sem befolyásolták a diákok akadémiai önértékelését. Bár a feltáró elemzésünk talált az elméletnek megfelelő evidenciákat – például, hogy egy gyengébben teljesítő padtárs növeli a diákok akadémiai önértékelését irodalom tantárgyból vagy hogy a lányok, ha rosszabbul teljesítő fiú padtárs mellett ülnek, akkor általában jobbnak gondolják saját képességeiket – azonban ezen eredmények egyike sem állta meg a helyét a többszörös hipotézisvizsgálat miatt alkalmazott mégoly visszafogott korrekció esetében sem. Végeredményben tehát az elmélettel ellentétben a randomizált terepkísérletünk nem szolgáltatott kísérleti bizonyítékot arra, hogy a kortársak hatással lennének a diákok akadémiai önértékelésére.

JEL: C93, I21, I24

Kulcsszavak: akadémiai önértékelés, kortársak hatása, társas összehasonlítás, Big-Fish-Little-Pond, hátrányos összehasonlítás, relatív depriváció, randomizált terepkísérlet, padtársak, Magyarország

Peer Effects on Academic Self-concept: A Large Randomized Field Experiment

Abstract

Social theories posit that peers affect students' academic self-concept (ASC). Most prominently, Big-Fish-Little-Pond, invidious comparison, and relative deprivation theories predict that exposure to academically stronger peers decreases students' ASC, and exposure to academically weaker peers increases students' ASC. These propositions have not yet been tested experimentally. We executed a large and pre-registered field experiment that randomized students to deskmates within 195 classrooms of 41 schools (N = 3,022). Our primary experimental analysis found no evidence of an effect of peer achievement on ASC in either direction. Exploratory analyses hinted at a subject-specific deskmate effect on ASC in verbal skills, and that sitting next to a lower-achieving boy increased girls' ASC (but not that sitting next to a higher-achieving boy decreased girls' ASC). Critically, however, none of these group-specific results held up to even modest corrections for multiple hypothesis testing. Contrary to theory, our randomized field experiment thus provides no evidence for an effect of peer achievement on students' ASC.

Keywords

Academic self-concept, peer effects, social comparison, Big-Fish-Little-Pond, invidious comparison, relative deprivation, randomized field experiment, deskmates, Hungary

1. Introduction

Academic self-concept (ASC) describes students' perception of their own academic ability (Shavelson, Hubner, and Stanton 1976). ASC matters because students invest in tasks in which they expect to succeed (Eccles et al. 1983), thus linking ASC to social and educational outcomes. For example, ASC correlates with student effort in homework and test preparation (Trautwein et al. 2009), academic motivation and achievement (Nagengast and Marsh 2012; Marsh and Martin 2011), college-major choices (Musu-Gillette et al. 2015), and the sorting of men into, and women out of, STEM careers (Oakes 1990; Seymour 1995; Nagy et al. 2006; Vinni-Laakso et al. 2019).

Building on relative deprivation and reference group theory (e.g., Merton 1968; Stouffer et al. 1949), sociologists have long argued that ASC is socially determined because students assess their own ability relative to that of their peers (Alwin and Otto 1977; Davis 1966; Drew and Astin 1972; Meyer 1970; Jonsson and Mood 2008; Rosenqvist 2018). This idea is most clearly articulated in Davis' (1966) foundational "frog-pond effect" and the modern "Big-Fish-Little-Pond" (BFLP) literature (Marsh 1987), echoing Veblen's earlier idea of "insidious comparisons" (Noe and Elifson 1975; Veblen 1899). They posit that exposure to higher achieving peers decreases students' ASC, and exposure to lower achieving peers increases students' ASC. Consonant with this tradition, sociologists have found that exposure to higher achieving peers is associated with lower college (Alwin and Otto 1977; Meyer 1970), and career ambitions (Davis 1966) and discourages advanced track choice in secondary school (Jonsson and Mood 2008; Rosenqvist 2018).

Prior research, however, is largely observational. This raises the question of whether the relationship between peer achievement and students' ASC is merely correlational and possibly an artifact of selection bias, or whether it represents a causal peer effect.

This article fills this gap by executing a large and pre-registered randomized field experiment of peer effects on students' ASC. Specifically, we randomized seating charts within 195 classrooms of 41 schools in Hungary for the duration of one whole semester (N = 3,022). Focusing on deskmates as close and intense peers, we tested whether a student's deskmate's baseline achievement, relative to the student's own baseline achievement, affects student's absolute, comparative, or subject-specific ASCs.

The results are surprising. Against the theoretically founded expectation that exposure to higher-achieving peers decreases and exposure to lower-achieving peers increases students' ASC, our primary pre-specified analyses find no evidence for a causal effect of peer achievement on ASC. This null result is robust to correction for measurement error. Furthermore, executing a large number of pre-registered exploratory models to probe for effect heterogeneity by school subject, gender, and other factors, we find very little; and the few estimates that reach conventional levels of statistical significance do not hold up to even modest corrections for multiple testing. In sum, our experiment provides no dependable evidence for a causal effect of peer achievement on students' ASC.

The article proceeds as follows. Section 2 outlines theoretical expectations. Section 3 describes the institutional setting and details the experimental design, data, and analysis. Section 4 presents empirical results. Section 5 concludes.

2. Literature review and theoretical expectations

Peer effects in education are a fertile ground for theory construction (Coleman et al. 1966; Hoxby and Weingarth 2005; Sacerdote 2011; 2014), and multiple mechanisms have been proposed to link peer exposures to ASC. Not all of these proposed mechanisms point in the same direction.

Prior work has primarily emphasized mechanisms suggested by social comparison and reference group theory (Festinger 1954; Merton 1968). The dominant expression of this tradition is Big-Fish-Little-Pond theory [BFLP] (Marsh and Parker 1984; Marsh 1987; Marsh et al. 2008), which posits that exposure to peers prompts students to engage in 'invidious comparisons' (Hoxby and Weingarth 2005; Veblen 1899): Occupying an inferior position relative to one's peers is said to initiate an upward comparison that depresses students' ASC by raising the reference point of good performance, causing ego-reduction. Conversely, occupying a superior position relative to one's peers would initiate a downward comparison that boosts students' ASC by lowering the reference point of good performance, causing ego-enhancement (Wayment and Taylor 1995; Gibbons, Benbow, and Gerrard 1994). In sum, BFLP theory predicts an inverse relationship between students' ASC and peers' achievement.

Others have argued that social comparison processes could also influence ASC in the opposite directions from those predicted by BFLP theory (Suls 2000; Suls, Martin, and Wheeler 2002). Instead of ego reduction, exposure to academically stronger peers might induce positive social comparisons(Marsh, Kong, and Hau 2000), whereby students identify with their successful peers and bask in their reflected glory (Burleson, Leach, and Harrington 2005; Collins 1996), thus raising students' ASC. Conversely, instead of ego enhancement, exposure to academically weaker peers might stoke fear of decline and thus decrease students' ASC (Wills 1981; Suls, Martin, and Wheeler 2002). This would result in a positive effect of peer achievement on students' ASC.

Yet other mechanism forego appeal to social comparisons altogether and link peer exposures to students' ASC via on students' learning and teachers' instruction. On the student side, exposure to high-achieving peers might promote students' own achievement through peer learning and hence boost their ASC. On the teacher side, a higher level of classroom achievement might lead teachers to raise expectations and quicken the pace of instruction (Duflo, Dupas, and Kremer 2011), thus leaving students of given ability behind and lowering their ASC.

Although the net effect of peer exposure on students' ASC via these disparate mechanisms is *a priori* ambiguous, prior evidence from observational studies is mostly consistent with the pattern predicted by BFLP theory: exposure to higher-achieving peers is associated with lower, and exposure to lower achieving peers is associated with higher ASC (e.g., Marsh and Yeung 1998; Marsh and Hau 2003; Seaton, Marsh, and Craven 2009; Loyalka, Zakharov, and Kuzmina 2018; see Fang et al. 2018 for a recent meta-analysis). Support also comes from a small number of recent quasi-experimental studies. For example, Elsner and Isphording (2017) exploit variation in the cohort composition of American high schools and find that students' rank within their grade level positively predicts their high school graduation and college enrollment, likely via increasing students' confidence and perceived intelligence. Pop-Eleches and Urquiola (2013) demonstrate that Romanian children who scored just above admission cutoffs for selective schools tend to perform worse through a reduction in confidence and/or self-esteem, potentially resulting from their lower relative ability compared to their classmates.

Prior evidence on peer effects on ASC, however, is limited in that it is largely observational. This raises the threat of selection bias and begs the question to what extent the observed correlations capture real causal effects. If, for example, students compare themselves downward in less selective schools and gain from this comparison, and compare themselves upward in more selective schools and suffer from this comparison, then school-average achievement would correlate negatively with students' ASC even if the former does not cause the latter (Dai and Rinn 2008). It is widely accepted that the best evidence for causal peer effects comes from randomized experiments, which rule out selection bias and other statistical artifacts by design (An 2018; Angrist 2014). Our study contributes the first large randomized field experiment of peer effects on ASC in a natural setting in order to probe causality. We focus on peer effects from students' deskmates because students (a) know their deskmates' academic achievement well and (b) use deskmates as a target of comparison, thus meeting key scope conditions of social comparison theory, as we document below.[1] Furthermore, deskmates are of interest since teachers around the world routinely assign seating charts, making deskmate assignments a promising target for largescale policy intervention.

Following the main thrust of the social comparison literature, we therefore pre-registered the primary hypothesis that exposure to a higher-achieving deskmate lowers, and exposure to a lower-achieving deskmate increases students' ASC.

Since prior research reports stronger negative correlations between peer achievement and students' ASC among older students and for verbal subjects, and weaker correlations in STEM

[1] Reviews of social comparison theory in general, and of comparison effects on ASC in particular, emphasize that social comparison is an active process in which students compare themselves to specific known peers (Gerber, Wheeler, and Suls 2018; Jansen, Boda, and Lorenz 2022). Comparison theorists have also long stressed the importance of physical proximity (Festinger, Schachter, and Back 1950). In part, such clarifications were offered in reaction to empirical work in the BFLP tradition that mostly studied ASC in relation to school-average peer achievement, which, critics claim, students are unlikely to know and unlikely to use as a target of comparison (Dai and Rinn 2008). We take no stand on stand on this debate and merely point out that deskmates are a plausible peer comparators.

fields and general skills (Fang et al. 2018; Kim, Liu, and Zhao 2022), we also pre-registered to investigate causal effect heterogeneity by gender, grade level, and school subject.

3. Institutional context, design, measures and methods

We studied peer effects on students' ASC by randomizing the seating charts in 195 3rd to 8th-grade classrooms of 41 rural Hungarian primary schools for the duration of the Fall semester 2017-18. Outcomes were collected through student surveys in the subsequent Spring semester 2018. The study was approved by the IRB offices at the Center for Social Sciences, Budapest, and the University of Wisconsin-Madison. Written consent was obtained at multiple points from school districts, school principals, teachers, and parents. A detailed pre-analysis plan was registered prior to the receipt of endline data. A replication package and the pre-analysis plan are available on the study's OSF page https://osf.io/gjxz7/ .

3.2. Institutional context: Schools and Deskmates

Primary education in Hungary starts at age 6 in 1st grade and ends with 8th grade. Primary schools are not tracked so that students across the ability spectrum are taught together in the same classroom. Students usually have a single teacher for all subjects from 1st through 4th grade and subject-specific teachers from 5th grade. Students form stable classrooms that advance from one grade level to the next together.

The core subjects in primary school are Hungarian literature, Hungarian grammar, and mathematics. Hungarian literature is a reading class, where students often read out loud. Hungarian grammar classes focus on spelling and writing. Depending on grade level, the three core subjects account for 7 and 10 hours of instruction per week, or between about one quarter and half of the average school day.[2]

Grades are determined by written exams, homework assignments, oral participation, and oral recitations. Written exams contribute the greatest weight in students' final grades and include frequent low-stakes teacher-written tests, administered once or twice per month, and high-stakes tests at the end of each semester in each core subject.

Students have many opportunities to glean the academic ability and achievement of their peers and likely know their deskmates' achievement better than that of any other classmate, as we argue with information from supplementary student and teacher surveys (see Appendix E for details): First, deskmates spend the longest time in closest proximity to each other among all school mates because seating charts are determined by homeroom teachers for the duration of a semester, and most subjects are taught in the same room. Second, teachers return written assignments and exams openly to students' desks, so that deskmates can see each other's grades. Third, deskmates routinely collaborate with each other on academic exercises and shared tasks. Teachers in Hungarian primary schools report that 61 percent of deskmates collaborate almost every lesson, and 95 percent of deskmates collaborate at least once a week, ensuring intimate familiarity with each other's academic performance even beyond official grades.

Proximity, duration of exposure, privileged access to grades, and detailed performance signals through dyadic collaboration put deskmates at an obvious information advantage with respect to each other's academic achievement and renders the hypothesis of peer effects from deskmates *a priori* plausible. Furthermore, students themselves testify to the salience of

---

[2] Data: http://eduline.hu/kozoktatas/2018/8/27/mit_tanulnak_a_diakok_az_iskolaban_98J42H and http://eduline.hu/kozoktatas/2018/8/28/felsos_kerettanterv_2018_Q86D33.

deskmates for academic comparisons, thus specifically substantiating the hypothesized mechanism from deskmate exposure via peer comparison to ASC. Specifically, among the 40 percent of primary school students who report that they compared their own performance to that of any peers in the classroom, 72 percent reported comparing their performance to their deskmate, exceeded only by the 87 percent who reported comparing their performance to their friends in the classroom (which may include deskmates). Importantly, more students say that they compare themselves to their deskmate than to the average student in the classroom (46 percent), which was the focus of much prior empirical research in the BFLP tradition.

3.2. Experimental Design

We recruited classrooms by contacting all primary schools in 7 contiguous counties of central and eastern Hungary in early 2017. In interested schools, we selected all 3rd through 8th grade classrooms that anticipated to (1) implement our randomized seating chart in at least three core subjects: Hungarian literature (reading), Hungarian grammar (writing), and mathematics (and in additional subjects if possible); (2) instruct all students in each of these subjects together in the same classroom (e.g., no ability grouping); and (3) maintain a grid-shaped classroom layout of free-standing front-facing desks that seat two students. Most participating schools were the only primary school in town. Participating schools are not nationally representative, having lower average test scores and parental education than the national average (see Appendix B for descriptives).

We randomized students within classrooms to freestanding, two-person, front-facing desks via unconstrained random partitioning. Randomization was based on the class rosters from the preceding spring 2017 semester. A replacement algorithm was stipulated to account for changes to class rosters via exits and entries during the summer.[3] We define the deskmate composition resulting from randomization and algorithm-based replacement as the intended seating chart.

Teachers were instructed to employ the intended seating chart in (at least) three subjects—mathematics, Hungarian literature, and Hungarian grammar—from the first day of classes (September 1, 2017) until the end of the fall semester (January 31, 2018). While teachers were permitted to reseat students after baseline for ethical reasons, we asked teachers to preserve the intended deskmate composition wherever possible.[4]

We pre-registered to exclude classrooms that did do not meet the various inclusion criteria, which resulted in an anticipated sample of 3,814 students at the time of pre-registration. Subsequent inspection of the data revealed some double entries, resulting in 3,803 unique cases. Five more students were excluded because their classroom was smaller than the pre-registered minimum class size of 10, and 36 students turned out to have left their classrooms before the intervention. As pre-registered, we further deleted 397 students who were randomized to sit alone. Among the remaining 3,365 students, 343 (10.19%) did not participate in the endline survey because they were absent on the day of the test or lacked parental consent. The final analytic sample thus contains N=3,022 students in 195 classrooms of 41 schools.

Compliance with the intended seating chart was high. Two weeks post baseline, teacher reports of the actual seating chart indicated that 94.2% of the students in the analytic sample sat

[3] 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.
[4] For example, if one student had to move to the front of the classroom for vision problems, we asked that her deskmate be moved with her.

next to their assigned deskmate. All analyses below are intent-to-treat analyses based on the intended seating chart.

3.3. Measures

3.3.1. Treatment variables

Our primary pre-registered treatment is the baseline GPA of student i's intended deskmate in classroom c and school s, defined as the average of deskmate's baseline grades[5] in the three core subjects of Hungarian literature, Hungarian grammar, and mathematics.[6] Each subject was graded on an integer scale from 1 (worst) to 5 (best). The mean GPA in the analytic sample was 3.71 (1.0 SD) (Table 1).

Table 1: Descriptive statistics for the analytic sample

	N (non-missing)	% Missing	Mean	SD	Min	Max
Outcome variables						
Average						
AASC	2,965	1.89	4.69	1.33	1	7
CASC	2,909	3.74	4.59	1.39	1	7
Subject-specific						
Grammar:						
AASC	2,842	5.96	4.61	1.45	1	7
CASC	2,721	9.96	4.58	1.50	1	7
Literature:						
AASC	2,874	4.90	4.90	1.53	1	7
CASC	2,750	9.00	4.73	1.54	1	7

[5] Baseline grades are teacher-reported and refer to the mid-term grades in the prior academic year (January 2017). We filled in missing teacher-reported grades with students' retrospectively self-reported end-of-year grades (3% of the cases).

[6]We also measure deskmate baseline ability using nationally standardized test scores in an un-preregistered robustness check (see results, below). These test scores became available for a subset of students only after the conclusion of the experiment. We consider GPA a more salient measure of peer exposure, since students are much more likely to know deskmates' grades than they are knowing one-time test scores.

Mathematics						
AASC	2,869	5.06	4.61	1.78	1	7
CASC	2,746	9.13	4.53	1.74	1	7
Treatment variables						
GPA						
DM Lower	2,908	3.77	0.34	0.47	0	1
DM Higher	2,908	3.77	0.33	0.47	0	1
Subject-specific grade						
Grammar:						
DM Lower	2,886	4.50	0.35	0.48	0	1
DM Higher	2,886	4.50	0.35	0.48	0	1
Literature						
DM Lower	2,888	4.43	0.34	0.47	0	1
DM Higher	2,888	4.43	0.34	0.47	0	1
Mathematics						
DM Lower	2,880	4.70	0.34	0.47	0	1
DM Higher	2,880	4.70	0.33	0.47	0	1
Control variables						
Own GPA	2,985	1.22	3.71	1.00	1	5
Behavior grade	2,884	4.57	4.30	0.82	2	5
Diligence grade	2,885	4.53	4.02	0.94	2	5
Girl	3,022	0	0.48	0.50	0	1
Age	3,022	0	11.88	1.82	8.2	17.5
Poor	2,873	4.93	0.10	0.30	0	1
Rich	2,873	4.93	0.10	0.30	0	1

Notes: N=3,022. DM: deskmate. GPA: Grade point average. AASC: Absolute academic self-concept. CASC: Comparative academic self-concept.

Our primary analysis divided deskmate's baseline GPA into three categories, $Higher_{ics}^{D}$, $Lower_{ics}^{D}$, or $Same_{ics}^{D}$, if the deskmate's baseline GPA was $\geq 2/3$ points higher, $\geq 2/3$ points lower, or within less than $\pm 2/3$ units of student i's own baseline GPA, respectively. This corresponds, for example, to sitting next to a deskmate who is better/worse by one grade in two out of the three subjects and is not worse/better in the other, or to sitting next to a deskmate who is better/worse by two grades in one subject and no worse/better in the others.

As secondary treatments, we analyzed the effects of deskmates' subject-specific grades in Hungarian literature, Hungarian grammar, and mathematics. We categorized deskmates' grades as being higher, lower, or the same if deskmate's grade in the subject was $\geq |1|$ grade higher, lower, or the same as student's own grade in the subject, respectively.

For exploratory analyses, we also divided deskmates' and students' own GPAs and grades into three categories based on classroom-specific quartiles, coded *[L]ow* (lowest quartile), *[M]iddle* (middle two quartiles), and *[H]igh* (highest quartile), and created nine product terms between deskmate's and student's own performance (LL_{ics}, LM_{ics}, LH_{ics}, ML_{ics}, MM_{ics}, ML_{ics}, HL_{ics}, HM_{ics}, HH_{ics}) to flexibly fit the joint distribution of students' and deskmates' baseline GPAs and grades.

3.3.2. Outcome variables

Our outcomes are students' absolute academic self-concept *(AASC)* and comparative academic self-concept *(CASC)*, which we measured using subject-specific items (Eccles et al. 1989; Eccles 1993; Musu-Gillette et al. 2015).[7] Outcomes were collected during a 45-minute teacher-administered in-class student survey at endline, one to ten weeks after completing the deskmate intervention. The survey instrument is available at the study's OSF page (blinded for review).

AASC evaluates students' academic self-concept without a reference point. It is measured separately for each of the three core subjects (Hungarian grammar, Hungarian literature, and mathematics) by asking, "In your opinion how good are you at [subject]?", coded from 1 ("I am very bad at [subject]") via 4 ("I am average at [subject]") to 7 ("I am very good at [subject]").

CASC evaluates students' academic self-concept relative to their classmates. It is measured separately for each of the three core subjects by asking: "Compared to your classmates how good are you at [subject]?", coded from 1 ("In the class, I am among the worst at [subject]")

[7] We also pre-registered to investigate subject liking (SL) as an affective analog to the cognitive dimension of ASC. Since SL did not add to our understanding of deskmate effects on ASC, we report results for the pre-registered analyses involving SL in Appendix C.

via 4 (“In the class, I am average at [subject]”) to 7 (“In the class, I am among the best at [subject]).”

Our primary outcomes, Y_{ics}, are students’ average AASC (mean 4.69, sd 1.33) and average CASC (mean 4.59, sd 1.39), respectively. We computed both outcomes as the average of each student’s non-missing responses across the three core subjects. Outcome distributions by students’ gender and baseline GPA are shown in Appendix Figures A1 and A2.

In exploratory analyses, we additionally analyzed subject-specific AASC and CASC in Hungarian literature, Hungarian grammar, and mathematics. We note that the average student considered themselves slightly “above average” in the absolute and comparative sense in all subjects (Table 1).

3.3.3. Control variables

As robustness checks and to improve efficiency, we control for students’ baseline characteristics, X_{ics}, measured before the start of the intervention. Control variables are mostly collected via teacher reports and administrative records. Classroom teachers reported student’s gender, age, a measure of socioeconomic status (SES) (“name the richest and poorest students in the classroom”), ethnicity (Roma, non-Roma), baseline GPA in the three core subjects (Hungarian literature, Hungarian grammar, mathematics), and grades in diligence and behavior, (ranging from 2 [worst] to 5 [best]). Table 1 presents summary statistics.

3.3.4. Missing values

Treatment variables and covariates had less than 5 percent missing values (Table 1). We imputed missing teacher reports on students’ baseline grades from student self-reports collected at endline. If missing values remained in some but not all baseline grades, we computed baseline GPA from non-missing grades. We dropped students with fully missing baseline grades (n = 37). We coded missing values on covariates as zero and included dummy variables controlling for missing status. We did not impute ethnicity (Roma, non-Roma), which was missing for 13 classrooms and was hence pre-registered not to be used in this study. We did not impute missing outcomes.

3.4. Analytic strategy

3.4.1. Primary analyses

Our primary pre-registered specification (Eq.1.) regressed student's average AASC or CASC, Y_{ics}, respectively, on two binary indicators for whether the student's deskmate had a lower ($Lower^{D}_{ics}$) or higher ($Higher^{D}_{ics}$) baseline GPA than the student; students' own GPA (GPA_{ics}) to control for confounding by the artifactual negative correlation between students' and deskmates' GPAs that is induced by randomizing students to desks; and classroom fixed effects (η_{cs}) to account for the experimental design, which randomized deskmates within classrooms:

$$Y_{ics} = a + b_1 Lower^{D}_{ics} + b_2 Higher^{D}_{ics} + b_3 GPA_{ics} + \eta_{cs} + e_{ics} \quad \text{(Eq. 1.)}$$

The coefficients b_1 and b_2 identify the causal effects of sitting next to an academically weaker or stronger deskmate, respectively, by virtue of randomizing the seating chart. The coefficient b_3 does not have a causal interpretation because students' own GPAs were not randomized. To improve statistical precision, we also estimated models that additionally controlled for students' baseline covariates (gender, age, SES [Rich and Poor], and baseline grades in behavior and diligence). We clustered standard errors at the school level.

Following convention, we conducted two-sided hypothesis tests for each coefficient, and we assessed statistical significance at the $\alpha = 0.05$ level. We also tested whether deskmate's relative standing explains any variation in academic self-concept by performing a joint F-test for $b_1 = b_2 = 0$ in each model. We did not penalize standard errors in this primary analysis.

3.4.2. Exploratory analyses

In addition to the primary analyses, we pre-registered several exploratory analyses. First, to explore whether deskmate effects on ASC varied by school subject, we analyzed subject-specific ASC by replacing the treatment variables in Eq.1. with indicators of whether the student's deskmate had a higher or lower baseline grade in Hungarian literature, Hungarian grammar, and mathematics, respectively.

Second, we explored whether deskmate effects on average academic self-concept varied by students' and deskmates' gender by estimating Eq.1. separately for male and female students and interacting deskmate's baseline achievement with deskmate's gender.

Third, we re-estimated the above models using more flexible specifications that allowed deskmate effects to vary freely with students' own baseline GPA. To this end, we regressed students' average and subject specific AASC and CASC on all 9 combinations of own and deskmate's GPA (categorized as low, middle, or high), treating middle students sitting next to middle deskmates as the reference category and classroom fixed effects.

$$Y_{ics} = a + b_1 LL_{ics} + b_2 LM_{ics} + b_3 LH_{ics} + b_4 ML_{ics} + b_5 MH_{ics} + b_6 HL_{ics} + b_7 HM_{ics} + b_8 HH_{ics} + \eta_{cs} + e_{ics} \quad \text{(Eq. 2.)}$$

Fourth, we explore whether deskmate effects varied by overall classroom performance or grade level (un-preregistered) by interacting deskmate GPA with (sample-centered) classroom-average GPA or with grade level.

Since multiple testing increases the risk of false positives (Type I errors), we pre-registered to penalize the statistical tests of our exploratory estimates using the procedure of Benjamini and Hochberg (1995), which holds the false-discovery rate at 5 percent of rejected null hypotheses within each set of exploratory models.[8] Following emerging conventions, we assess statistical significance first according to unpenalized traditional standards and second according to Benjamini and Hochberg's criterion.

4. Results

We executed all pre-registered analyses. Pre-registered analyses not shown here are shown in the Appendix.

[8] The literature has not yet established firm conventions for how to group analyses into sets within which the Benjamini-Hochberg correction should control the false-discovery rate. We consider relatively narrow sets of tests for the coefficients on key exposure (deskmate) variables within each of our four sets of exploratory analyses. This amounts to small corrections relative to traditional, uncorrected, standard errors. Since these corrections fail to detect statistically significant effects, more conservative corrections that consider larger groups of exploratory tests would necessarily also fail to detect statistically significant effects.

4.1 Balance checks

The key advantage of randomized experiments is that they warrant causal inferences by creating comparable ("balanced") treatment and control groups. In order to assess the success of randomization, we conducted balance checks following Guryan, Kroft, and Notowidigdo (2009). These checks separately regress each baseline characteristic on the corresponding baseline characteristic of students' deskmate's, the leave-one-out mean characteristic in the classroom and classroom fixed effects; with standard errors clustered at the school level. This procedure circumvents the artifactual correlation between own and deskmates' characteristics that is induced by randomizing students to desks within classrooms. Using this approach, we found no substantively meaningful or statistically significant association between any of the students' and their deskmates' baseline characteristics, which indicates excellent balance and hence successful randomization (Appendix Table A1).

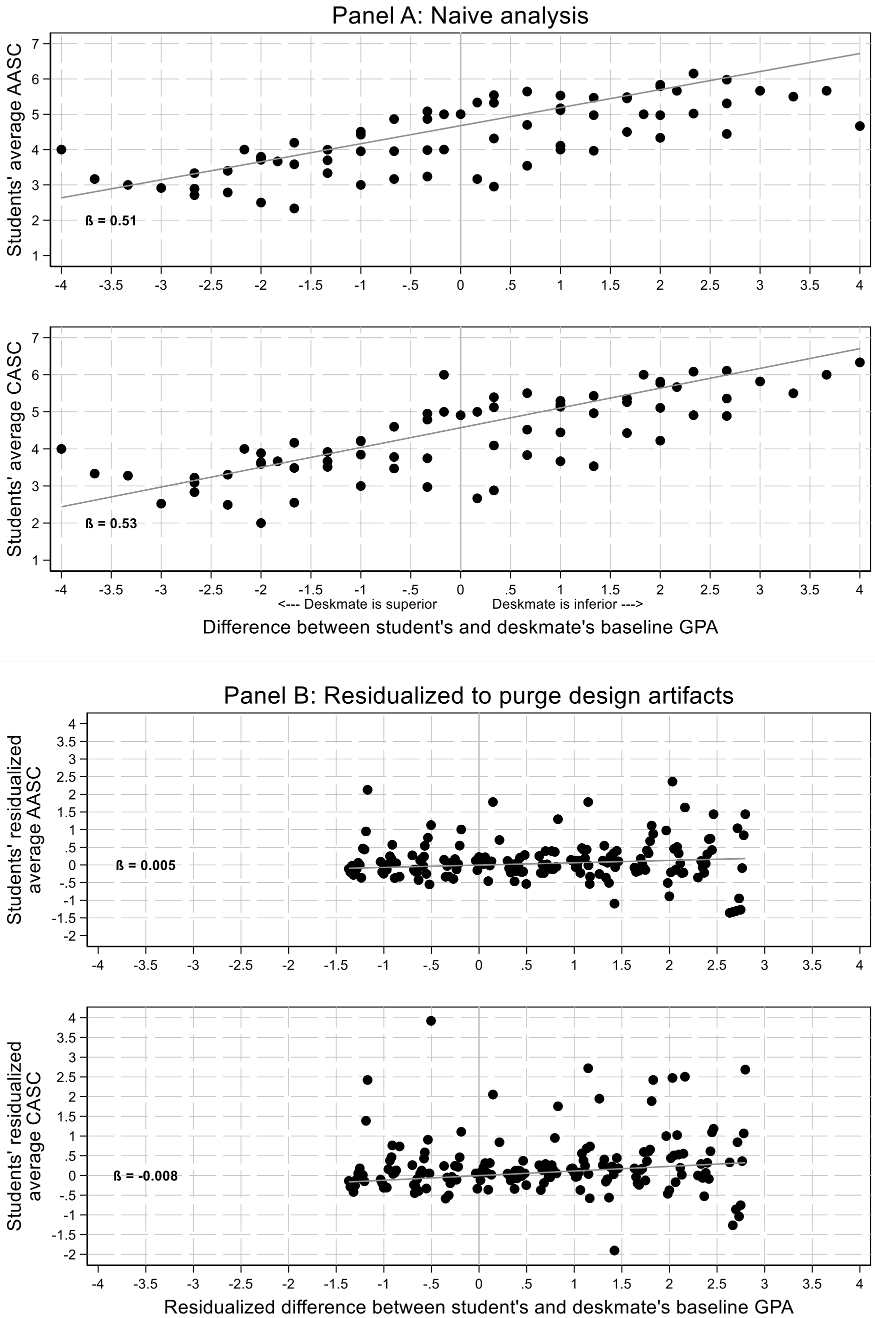
4.2. Peer correlations and their artifacts

We begin our analysis by demonstrating how a naïve analysis—even of a randomized experiment—can falsely suggest peer effects where none exist. Panel A of Figure 1 shows our raw data, which exhibit a strong positive association between students' ASC and sitting next to an academically weaker deskmate. For each one-grade-point decrease in deskmate's baseline GPA relative to the student's own baseline GPA, students' average AASC and CASC increase by half a point on a 7-point scale.

While these associations appear to align with the peer effects predicted by BFLP and invidious comparison theories, they are, in fact, spurious and must not be interpreted as causal peer effects, for two reasons. First, this naïve analysis captures both within- and between-classroom variation and hence neglects students' selection into classrooms (recall that we randomize deskmates within, but not across, classrooms). Second, the naïve analysis neglects that randomizing students within larger pools (here, classrooms, but the same would be true if randomizing students within grades or schools) necessarily induces a negative association between students' and their peers' baseline characteristics (Angrist 2014). For example, the weakest student necessarily sits next to a stronger student, and vice versa. Since students' own GPA is hence positively associated with sitting next to a weaker deskmate, and is also positively associated with students' own ASC, failure to control for own GPA will induce an artifactual positive association between sitting next to a weaker deskmate and students own ASC, just as seen in Figure 1, Panel A.

Panel B presents the proper experimental specification, which eliminates the statistical artefacts of the naïve analysis by controlling (residualizing) for classroom-fixed effects (to account for randomization within classrooms) and students' own baseline GPA (to account for the negative association between students' and deskmates' baseline GPAs induced by randomization). The residualized association is clearly null, providing no evidence of a causal effect of deskmates' baseline GPA on students' ASC. Next, we confirm this conclusion by reporting results for our pre-registered analyses.

Figure 1. Association between students' academic self-concept (average AASC and CASC) and the difference between student's and deskmate's baseline GPA (a) in the raw data, and (b) residualized to purge design artifacts. Binned Scatter Plots.



Notes: $N = 3{,}022$. GPA: Grade point average. AASC: Absolute academic self-concept. CASC: Comparative academic self-concept. Panel B residualizes the outcome (average ASC) and the exposure (difference in baseline GPAs) with respect to students' own baseline GPA (to remove the artifactual correlation between outcome and exposure unduced by random assignment of students to desks within classrooms) and classroom fixed effects (to account for randomization within classrooms). β is the coefficient on exposure in an OLS regression of outcome on exposure (unadjusted in Panel A, residualized in Panel B).

4.3. Effects of deskmate achievement on students' average academic self-concept

Table 2 reports our primary confirmatory estimates for the causal effect of deskmates' baseline GPA on students' average AASC and average CASC across the three core subjects of Hungarian grammar, literature, and mathematics. Columns 1-2 report models without baseline controls, and columns 3-4 report models with baseline controls. All models include classroom-fixed effects (Full regression tables are shown in in Appendix Table A2).

Table 2: Estimated causal effects of deskmate baseline GPA on students' average academic self-concepts (average AASC & CASC), pre-registered primary analysis

	Without controls		With controls	
	(1) AASC	(2) CASC	(3)$ AASC	(4) $ CASC
DM Lower	0.03	0.06	0.03	0.05
	(0.04)	(0.05)	(0.04)	(0.05)
	[0.02]	[0.05]	[0.02]	[0.04]
DM Higher	-0.05	-0.01	-0.04	0.00
	(0.06)	(0.06)	(0.06)	(0.05)
	[-0.04]	[-0.01]	[-0.03]	[0.00]
Own GPA	1.01**	1.08**	0.90**	0.95**
	(0.03)	(0.04)	(0.05)	(0.06)
Constant	Yes	Yes	Yes	Yes
Mean of the dependent variable	4.692	4.589	4.692	4.589
Controls	No	No	Yes	Yes
Observations	2,856	2,805	2,856	2,805
R-squared	0.60	0.58	0.61	0.58
F (DM Lower = DM Higher = 0	0.94	1.15	0.64	0.73
Two-sided p-value	0.40	0.33	0.53	0.37
F (DM Lower = DM Higher	1.74	1.59	1.16	0.81
Two-sided p-value	0.19	0.21	0.29	0.49

Note: AASC is the absolute academic self-concept. CASC is the comparative academic self-concept.
All models control for classroom fixed effects. Standard errors are clustered at the school level.
Robust standard errors in parentheses. ** p<0.01, * p<0.05, + p<0.1
Cohen's D effect sizes (the coefficient divided by the standard deviation) are in [sqaured breackets]
Columns 1-2 report models without control variables, and columns 3-4 report models that control for baseline control variables.
Control variables: Girl (=1), Age, Poor (=1), Rich (=1), Baseline behavior grade in dummies (grade 5 is the reference category), Baseline diligence grade in dummies (grade 5 is the reference category). We code missing values in the covariates as zero and enter dummy variables to control for missingness. Missingness in the variables Rich and Poor are controlled by classroom fixed effects, as missingness in these variables affects entire classrooms.
$ Regression tables for all control variables are shown in Appendix Table A2.

We find no statistically significant evidence that sitting next to a higher-performing deskmate decreases (or otherwise affects) students' average AASC or CASC at the conventional 5 percent level of statistical significance. We also find no statistically significant evidence that sitting next to a lower-performing deskmate increases (or otherwise affects) these outcomes. Although the point estimates for AASC and CASC largely point in the expected direction (positive for sitting next to a lower-, and negative for sitting next to a higher-performing deskmate), the effect sizes are substantively very small. For example, sitting next to a lower-performing deskmate is estimated to increase students' own average AASC by only 0.02 standard deviations (Column 1). Since standard errors are tight, these null results are not the consequence of imprecise estimation. For both outcomes, non-significant F-tests (for joint-insignificance of the deskmate indicators) document that deskmates' achievement does not cause variation in students' average AASC and CASC.

To evaluate whether these null results are due to measurement error in the independent variable (which might attenuate the estimates toward zero), we repeated the main analysis for the subset of $N = 612$ 7^{th} and 8^{th} grade students for whom we were able to obtain nationally standardized test scores from Hungary's comprehensive National Assessment of Basic Competencies (NABC), measured in 6^{th} grade. This analysis was not pre-registered. Using NABC scores rather than GPA to measure students' and deskmates' baseline achievement, we still find no evidence of deskmate effects on students' average AASC or CASC either (results shown in Appendix D. This suggests that the null results of our pre-registered analysis are not due to measurement error.

4.4. Effects of deskmate achievement on subject-specific academic self-concept

Since the null results reported in Table 2 average deskmate exposure and student outcomes across the three core school subjects, we next explored the effect of deskmate's baseline grades on students' subject-specific ASC, relative to students' own grades in that subject. Table 3 shows estimates for the subject-specific effects of deskmates' relative baseline grades in grammar, literature, and mathematics, respectively, on students' subject-specific AASC and CASC, respectively.

Table 3: Estimated causal effects of deskmates' subject-specific baseline grades on students' academic self-concepts (ASC & CASC) in the same subject, by school subject (pre-registered exploratory analyses)

		Without controls		With controls	
		(1) AASC	(2) CASC	(3) AASC	(4) CASC
Panel A: Grammar	DM Lower in grammar	0.11+ (0.06) [0.08]	0.11 (0.08) [0.07]	0.12* (0.06) [0.08]	0.12 (0.08) [0.08]
	DM Higher in grammar	0.03 (0.07) [0.02]	0.04 (0.07) [0.03]	0.05 (0.07) [0.03]	0.07 (0.07) [0.05]
	Own Grammar grade	0.84** (0.04)	0.92** (0.05)	0.56** (0.07)	0.59** (0.08)
	Constant	Yes	Yes	Yes	Yes
	Mean of the dependent variable	4.615	4.588	4.615	4.588
	Controls	No	No	Yes	Yes
	Observations	2,723	2,610	2,723	2,610
	R-squared	0.45	0.44	0.48	0.48
	F(DMLow=DMHigh=0)	2.14	1.02	2.62	1.37
	Two-sided p-value	0.13	0.37	0.08	0.51
	F (DM Low = DM High)	1.78	0.67	1.35	0.44
	Two-sided p-value	0.19	0.42	0.25	0.27

Panel B: Literature	DM Lower in literature	-0.06 (0.07) [-0.04]	-0.00 (0.08) [-0.00]	-0.06 (0.07) [-0.04]	0.00 (0.07) [0.00]
	DM Higher in literature	-0.19* (0.08) [-0.12]	-0.16* (0.07) [-0.11]	-0.16* (0.07) [-0.11]	-0.14* (0.07) [-0.09]
	Own Literature grade	0.71** (0.03)	0.79** (0.04)	0.39** (0.05)	0.48** (0.05)
	Constant	Yes	Yes	Yes	Yes
	Mean of the dependent variable	4.902	4.725	4.902	4.725
	Controls	No	No	Yes	Yes
	Observations	2,749	2,634	2,749	2,634
	R-squared	0.38	0.41	0.42	0.45
	F(DMLow=DMHigh=0)	3.05	3.81	2.60	2.78
	Two-sided p-value	0.06	0.03	0.09	0.07
	F (DM Low = DM High)	2.78	5.39	2.11	4.12
	Two-sided p-value	0.10	0.03	0.15	0.05
Panel C: Mathematics	DM Lower in mathematics	-0.01 (0.07) [-0.01]	-0.02 (0.06) [-0.01]	-0.01 (0.07) [-0.01]	-0.02 (0.06) [-0.01]
	DM Higher in mathematics	-0.07 (0.09) [-0.04]	0.01 (0.10) [0.01]	-0.02 (0.10) [-0.01]	0.07 (0.10) [0.04]
	Own Mathematics grade	0.93** (0.06)	0.98** (0.06)	0.72** (0.08)	0.73** (0.07)
	Constant	Yes	Yes	Yes	Yes
	Mean of the dependent variable	4.616	4.531	4.616	4.531
	Controls	No	No	Yes	Yes
	Observations	2,745	2,632	2,745	2,632
	R-squared	0.41	0.40	0.44	0.44
	F(DMLow=DMHigh=0)	0.28	0.07	0.03	0.44
	Two-sided p-value	0.75	0.77	0.97	0.36
	F (DM Low = DM High)	0.36	0.08	0.01	0.86
	Two-sided p-value	0.55	0.93	0.94	0.65

Note: Each panel is estimated separately.
AASC is the absolute academic self-concept. CASC is the comparative academic self-concept.
All models control for classroom fixed effects. Standard errors are clustered at the school level.
Robust standard errors in parentheses. ** p<0.01, * p<0.05, + p<0.1
Cohen's D effect sizes (the coefficient divided by the standard deviation) are in [sqaured breackets]
Columns 1-2 report models without control variables, and columns 3-4 report models that control for baseline control variables.
Control variables: Girl (=1), Age, Poor (=1), Rich (=1), Baseline behavior grade in dummies (grade 5 is the reference category), Baseline diligence grade in dummies (grade 5 is the reference category). We code missing values in the covariates as zero and enter dummy variables to control for missingness. Missingness in the variables Rich and Poor are controlled by classroom fixed effects, as missingness in these variables affects entire classrooms.
No deskmate coefficient remains statistically significant after Benjamini-Hochberg correction for multiple testing.

Results lend at best weak and inconsistent support to the hypotheses that a higher-achiving deskmate decreases, and a lower-achiving deskmate increases, students' ASC.

In grammar (where instruction centers on writing), in line with expectations, sitting next to a deskmate with a lower baseline grammar grade increases students' AASC for grammar by about 0.1 units on the five-point grading scale (0.08 of a standard deviation). But this estimate reaches the conventional, uncorrected, 5-percent level of statistical significance only after controlling for baseline covariates.Contrary to expectation, sitting next to a deskmate with a higher baseline grade in grammar, also appears to increase (not decrease) students' AASC and CASC in grammar (not statistically significant).

In literature (where instruction centers on reading), in line with expectations, sitting next to a deskmate with a higher literature-grade decreases students' AASC and CASC by about 0.2 units on a five-point scale (0.1 standard deviations). These estimates are significant at the conventional, uncorrected, 5 percent level for AASC and CASC. Contrary to expectation, however, sitting next to a deskmate with a lower baseline grade in literature, also appears to decrease (not increase) students' AASC and CASC in literature (not significant).

We find no statistically significant signal in either direction for sitting next to a deskmate with a higher or lower baseline grade in mathematics, respectively, with or without controls.

In sum, these estimates do not consistently support the hypothesis that sitting next to an academically stronger deskmate decreases, and sitting next to an academically weaker deskmate increases, students ASC in any subject.

Critically, none of the subject-specific estimates remain statistically significant after adjusting for multiple testing (Benjamini and Hochberg 1995). This indicates that the few conventionally statistically significant coefficients in the tests that do not control the false-discovery rate for multiple testing may simply be due to chance.[9]

4.5. Effects of deskmate achievement on average academic self-concept by gender

Our pre-registered gender-specific analyses similarly reveal at best weak evidence for deskmate effects on students' ASC. Boys' ASC does not meaningfully respond to deskmate baseline GPA, regardless of whether the deskmate is a boy or a girl (point estimates are small and statistically insignificant; Table 4, Panel A). Similarly, girls' ASC does not respond to deskmates' GPA if the deskmate is a girl (Table 4, Panel B). Girls' AASC and CASC at first do appear to increase by about 0.1 SD when sitting next to an academically weaker boy. This estimate, however, no longer reaches statistical significance when it is adjusted for multiple testing. There is no evidence that girls' ASC responds to sitting next to an academically stronger boy.

Table 4: Gender differences in the causal effects of deskmate baseline average GPA on students' average ASC by students' and deskmate's gender (pre-registered exploratory analyses)

			Without controls		With controls	
			(1) AASC	(2) CASC	(3) AASC	(4) CASC
Panel A: Boy	DM is male	DM Lower	-0.09	-0.06	-0.10	-0.08
			(0.07)	(0.10)	(0.07)	(0.10)
		DM Higher	-0.08	-0.13	-0.09	-0.14
			(0.10)	(0.10)	(0.10)	(0.10)
	DM is female	DM Lower	-0.07	0.03	-0.08	0.03
			(0.10)	(0.11)	(0.10)	(0.11)
		DM Higher	-0.02	0.05	-0.00	0.08
			(0.10)	(0.08)	(0.10)	(0.08)
	Observations		1,480	1,456	1,480	1,456
	R-squared		0.64	0.60	0.65	0.60

[9] Here, we performed corrections for multiple testing separately for the deskmate coefficients in each column of Table 3. Hence, considering all deskmate coefficients in Table 3 together would also not yield significant results.

Panel A: Girl	DM is male	DM Lower	0.18*	0.21*	0.17*	0.20*
			(0.08)	(0.10)	(0.08)	(0.10)
		DM Higher	-0.00	-0.02	0.01	0.01
			(0.11)	(0.11)	(0.10)	(0.09)
	DM is female	DM Lower	0.09	0.07	0.08	0.06
			(0.08)	(0.09)	(0.08)	(0.08)
		DM Higher	-0.10	-0.02	-0.09	-0.01
			(0.14)	(0.12)	(0.13)	(0.12)
	Observations		1,376	1,349	1,376	1,349
	R-squared		0.65	0.63	0.65	0.65

Note: AASC is the absolute academic self-concept. CASC is the comparative academic self-concept.
All models control for classroom fixed effects, and contains the follwong variables: DM Girl; Own GPA; DM Girl × Own GPA, and Constant.
Standard errors (in parentheses) are clustered at school level. ** $p<0.01$, * $p<0.05$, + $p<0.1$.
Columns 1-2 report models without control variables, and columns 3-4 report models that control for baseline control variables.
Control variables: Girl (=1), Age, Poor (=1), Rich (=1), Baseline behavior grade in dummies (grade 5 is the reference category), Baseline diligence grade in dummies (grade 5 is the reference category). We code missing values in the covariates as zero and enter dummy variables to control for missingness. Missingness in the variables Rich and Poor are controlled by classroom fixed effects, as missingness in these variables affects entire classrooms.

Appendix Tables A2 and A3 show additional gender-specific analyses, none of which show statistically significant results after corretions for multiple hypothesis testing.

4.6. More Flexible Specifications

We conducted numerous additional heterogeneity analyses. First, as pre-registered, we explored whether the effect of the deskmate's relative achievement on students' academic self-concept varied by students' own baseline achievement (Eq. 2) in the above analyses. Results did not reveal systematic patterns; and the very small number of estimates that were statistically significant at conventional, uncorrected levels of statistical significance did not remain significant after corrections for multiple testing, suggesting chance associations (Appendix Figure A3).

Second, we explored other aspects of heterogeneity by grade-level (Table A5, un-preregistered) and classroom-average GPA (Table A6, pre-registered). None of these analyses yielded patterns of results that survived even modest corrections for multiple testing.

5. Discussion

Several social theories posit that peers influence students' academic self-concept (ASC). Most prominently, BFLP theory and invidious comparison models in the tradition of relative deprivation theory predict that exposure to academically stronger peers depresses, and exposure to academically weaker peers increases students' academic self-concept. Supportive evidence for peer effects on ASC, however, is largely observational and correlational.

We executed the first large randomized field experiment of peer effects on ASC by randomizing the seating charts of 195 3rd-8th grade classrooms. Although a naïve, observational, analysis found the familiar positive association between students' ASC and exposure to a relatively weaker peer in line with theoretical predictions (Figure 1A), our well-specified experimental analysis revealed this association to be a statistical artifact (Figure 1B).

Our primary pre-registered analysis found no positive or negative causal effects of exposure to higher- or lower-achieving deskmates on students' average ASC, regardless of whether ASC was measured on an absolute or on a relative scale. We further found no statistically dependable evidence for heterogeneous peer effects on ASC by school subject, students' own baseline grades, gender, or any other investigated characteristic. Most point estimates were substantively small and did not align with the patterns predicted by BFLP theory or gendered variants of social comparison theory. What few conventionally statistically significant estimates we found across our many analyses were not robust to even modest corrections for multiple hypothesis testing. Our experiment therefore provides no dependable evidence that exposure to close peers of differing baseline achievement levels within the classroom affects students' ASC on average or in any subgroup of students.

Our null findings have implications for policy and theory. For policy, they suggest that intervening on students' close-peer environment does not affect students' ASC, at least in our setting. Even if ASC affects downstream outcomes, intervention on students' close-peer environment hence is not a promising policy lever for promoting desired outcomes via the ASC mechanism.

For theory, our null findings fail to support the predictions of any theory of peer effects on ASC. This is compatible with at least three different interpretations. First, peer exposures may not initiate the invidious comparison process that is often hypothesized to connect peer exposures to ASC (Marsh et al. 2008). This possibility is supported by our out-of-sample survey evidence that students simply do not engage in much (conscious) comparison: although most Hungarian students who compare themselves to classmates compare themselves to their deskmates, less than half of all students report comparing themselves to any classmate.

Second, even if peer exposures initiate invidious comparisons, invidious comparisons may not meaningfully affect ASC, perhaps because ASC is primarily driven by students' own achievement.

Third, even if peer exposures initiate invidious comparisons, and invidious comparisons affect ASC, peer exposures may additionally initiate competing mechanisms that cancel out the effect of invidious comparisons. For example, while some students' ASC may diminish from invidious comparison to an academically stronger peer, other students' ASC may increase because they identify with the stronger peer ("reflected glory") (Marsh et al. 2008 Marsh, Kong, and Hau 2000; Sacerdote 2014). Unfortunately, our experiment only identifies the net effect of peer exposure on ASC and cannot disambiguate between these candidate explanations for our null findings.

It merits emphasizing that our experiment, narrowly interpreted, only informs the effect (or lack thereof) of sitting next to a deskmate of a given baseline achievement on students' ASC. Adopting the language of BFLP theory, our "pond" is the desk—a "little pond," indeed. Furthermore, our results may pertain only to our institutional setting, Hungarian primary schools. Hence, we cannot rule out that the ability distributions of larger peer environments, such as classrooms, grades, or schools, affect students' ASC, nor that deskmates may affect ASC in other institutional environments.

For example, it is possible that larger peer environments may activate new mechanisms that are not available at the desk level. Specifically, increasing the ability of the average peer in the classroom or school (e.g., via ability grouping or tracking) may affect teachers' teaching style and their expectations of students (Duflo, Dupas, and Kremer 2011): as classroom ability increases, teachers may increase expectations, such that a student of given ability may feel increasingly ill-matched to the task. By contrast, perturbing the seating chart is unlikely to affect teachers' teaching styles and expectations, since teachers are unlikely to calibrate curricular standards and performance expectations at the desk level.

That said, our failure to detect causal effects of peer ability on students' ASC at the desk level in a large pre-registered randomized field experiment raises the urgency of investigating other peer effects on ASC with similarly dependable research designs also at larger levels of peer exposure and in other institutional settings. If theory and policy depend on whether peer composition affects students' educational trajectories and social stratification via ASC, then more evidence to substantiate the existence of these peer effects is needed.

References

Alwin, Duane F., and Luther B. Otto. 1977. "High School Context Effects on Aspirations." *Sociology of Education* 50 (4): 259–73. https://doi.org/10.2307/2112499.

An, Weihua. 2018. "Causal Inference with Networked Treatment Diffusion." *Sociological Methodology*. https://doi.org/10.1177/0081175018785216.

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

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.

Burleson, Kathryn, Colin Wayne Leach, and David M. Harrington. 2005. "Upward Social Comparison and Self-Concept: Inspiration and Inferiority among Art Students in an Advanced Programme." *British Journal of Social Psychology* 44 (1): 109–23. https://doi.org/10.1348/014466604X23509.

Coleman, James S., E.Q Campbell, C.J. Hobson, F. McPartland, A.M. Mood, and F.D Weinfeld. 1966. *Equality of Educational Opportunity*. *Socio-Economic Planning Sciences*. Vol. 2. U.S. Government Printing Office. https://doi.org/10.1016/0038-0121(69)90029-9.

Collins, Rebecca L. 1996. "For Better or Worse：The Impact of Upward Social Comparison." *Psychological Bulletin* 119 (1): 51–69.

Dai, David Yun, and Anne N. Rinn. 2008. "The Big-Fish-Little-Pond Effect: What Do We Know and Where Do We Go from Here?" *Educational Psychology Review* 20 (3): 283–317. https://doi.org/10.1007/s10648-008-9071-x.

Davis, James A. 1966. "The Campus as a Frog Pond: An Application of the Theory of Relative Deprivation to Career Decisions of College Men." *American Journal of Sociology* 72 (1): 17–31. https://doi.org/10.1086/224257.

Drew, David E., and Alexander W. Astin. 1972. "Undergraduate Aspirations: A Test of Several Theories." *American Journal of Sociology* 77 (6): 1151–64. https://doi.org/10.1086/225262.

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.

Eccles, Jacquelynne S., T. F. Adler, R. Futterman, S. B. Goff, C. M. Kaczala, J. Meece, and C. Midgley. 1983. "Expectancies, Values and Academic Behaviors." In *Achievement and Achievement Motives*, edited by J. T. Spence. San Francisco: W. H. Freeman.

Eccles, Jacquelynne S., Allan Wigfield, Constance A. Flanagan, Chrity Miller, David A Reuman, and Dpris Yee. 1989. "Self- Concepts, Domain Values, and Self-Esteem: Relations and Changes at Early Adolescence." *Journal of Personality* 57 (2): 283–310. https://doi.org/http://dx.doi.org/10.1111/j.1467-6494.1989.tb00484.x.

Elsner, Benjamin, and Ingo E Isphording. 2017. "A Big Fish in a Small Pond: Ability Rank and Human Capital Investment." *Journal of Labor Economics* 35 (3): 787–828. https://doi.org/10.1086/690714.

Fang, Junyan, Xitong Huang, Minqiang Zhang, Feifei Huang, Zhe Li, and Qiting Yuan. 2018. "The Big-Fish-Little-Pond Effect on Academic Self-Concept: A Meta-Analysis." *Frontiers in Psychology* 9 (AUG): 1–11. https://doi.org/10.3389/fpsyg.2018.01569.

Festinger, Leon. 1954. "A Theory of Social Comparison Processes." *Human Relations* 7 (2): 117–40. https://doi.org/10.1177/001872675400700202.

Festinger, Leon, Stanley Schachter, and Kurt Back. 1950. *Social Pressures in Informal Groups; a Study of Human Factors in Housing*. Oxford, England: Harper Collins Publishers.
Gerber, J. P., Ladd Wheeler, and Jerry Suls. 2018. "A Social Comparison Theory Meta-Analysis 60+ Years On." *Psychological Bulletin* 144 (2): 177–97. https://doi.org/10.1037/bul0000127.
Gibbons, Frederick X., Camilla Persson Benbow, and Meg Gerrard. 1994. "From Top Dog to Bottom Half: Social Comparison Strategies in Response to Poor Performance." *Journal of Personality and Social Psychology* 67 (4): 638–52. https://doi.org/10.1037/0022-3514.67.4.638.
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.
Hoxby, Caroline M, and Gretchen Weingarth. 2005. "Taking Race out of the Equation: School Reassignment and the Structure of Peer Effects." *Working Paper*. https://doi.org/10.1.1.75.4661.
Jansen, Malte, Zsófia Boda, and Georg Lorenz. 2022. "Social Comparison Effects on Academic Self-Concepts—Which Peers Matter Most?" *Developmental Psychology*. https://doi.org/10.1037/dev0001368.
Jonsson, Jan O., and C. Mood. 2008. "Choice by Contrast in Swedish Schools: How Peers' Achievement Affects Educational Choice." *Social Forces* 87 (2): 741–65. https://doi.org/10.1353/sof.0.0135.
Kelley, Harold H. 1952. "Two Functions of Reference Groups." In *Readings in Social Psychology*, edited by G. E. Swanson, T. M. Newcomb, and E. L. Hartley, 410–14. New York: Holt, Rinehart & Winston.
Kim, Jinho, Ran Liu, and Xiaohang Zhao. 2022. "A Big (Male) Fish in a Small Pond? The Gendered Effect of Relative Ability on STEM Aspirations under Stereotype Threat." *European Sociological Review*, October. https://doi.org/10.1093/esr/jcac037.
Loyalka, Prashant, Andrey Zakharov, and Yulia Kuzmina. 2018. "Catching the Big Fish in the Little Pond Effect: Evidence from 33 Countries and Regions." *Comparative Education Review* 62 (4): 542–64. https://doi.org/10.1086/699672.
Marsh, Herbert W. 1987. "The Big-Fish-Little-Pond Effect on Academic Self-Concept." *Journal of Educational Psychology* 79 (3): 280–95. https://doi.org/10.1037/0022-0663.79.3.280.
Marsh, Herbert W., and Kit-Tai Hau. 2003. "Big-Fish--Little-Pond Effect on Academic Self-Concept: A Cross-Cultural (26-Country) Test of the Negative Effects of Academically Selective Schools." *American Psychologist* 58 (5): 364–76. https://doi.org/10.1037/0003-066X.58.5.364.
Marsh, Herbert W., Chit Kwong Kong, and Kit Tai Hau. 2000. "Longitudinal Multilevel Models of the Big-Fish-Little-Pond Effect on Academic Self-Concept: Counterbalancing Contrast and Reflected-Glory Effects in Hong Kong Schools." *Journal of Personality and Social Psychology* 78 (2): 337–49. https://doi.org/10.1037/0022-3514.78.2.337.
Marsh, Herbert W., and Andrew J Martin. 2011. "Academic Self-Concept and Academic Achievement: Relations and Causal Ordering." *The British Journal of Educational Psychology* 81 (Pt 1): 59–77. https://doi.org/10.1348/000709910X503501.
Marsh, Herbert W., Marjorie Seaton, Ulrich Trautwein, Oliver Lüdtke, K. T. Hau, Alison J. O'Mara, and Rhonda G. Craven. 2008. "The Big-Fish–Little-Pond-Effect Stands Up to

Critical Scrutiny: Implications for Theory, Methodology, and Future Research." *Educational Psychology Review* 20 (3): 319–50. https://doi.org/10.1007/s10648-008-9075-6.
Marsh, Herbert W., and Alexander Seeshing Yeung. 1998. "Longitudinal Structural Equation Models of Academic Self-Concept and Achievement: Gender Differences in the Development of Math and English Constructs." *American Educational Research Journal* 35 (4): 705–38. https://doi.org/10.3102/00028312035004705.
Merton, Robert K. 1968. *Social Theory and Social Structure*. New York: The Free Press.
Meyer, John W. 1970. "High School Effects on College Intentions." *American Journal of Sociology* 76 (1): 59–70. https://doi.org/10.1086/224906.
Musu-Gillette, Lauren E., Allan Wigfield, Jeffrey R. Harring, and Jacquelynne S. Eccles. 2015. "Trajectories of Change in Students' Self-Concepts of Ability and Values in Math and College Major Choice." *Educational Research and Evaluation* 21 (4): 343–370. https://doi.org/10.1080/13803611.2015.1057161.
Nagengast, Benjamin, and Herbert W. Marsh. 2012. "Big Fish in Little Ponds Aspire More: Mediation and Cross-Cultural Generalizability of School-Average Ability Effects on Self-Concept and Career Aspirations in Science." *Journal of Educational Psychology* 104 (4): 1033–53. https://doi.org/10.1037/a0027697.
Noe, Francis P., and Kirk W. Elifson. 1975. "An 'Invidious Comparison,' Class and Status, 1929-60: Effects of Employment, Cost and Time on Veblen's Theory of Class." *American Journal of Economics and Sociology* 34 (4): 381–96. https://doi.org/10.1111/j.1536-7150.1975.tb01200.x.
Oakes, Jeannie. 1990. "Opportunities, Achievement, and Choice: Women and Minority Students in Science and Mathematics." *Review of Research in Education* 16 (1): 153–222. https://doi.org/10.3102/0091732X016001153.
Pop-Eleches, Cristian, and Miguel Urquiola. 2013. "Going to a Better School: Effects and Behavioral Responses." *American Economic Review* 103 (4): 1289–1324. https://doi.org/10.1257/aer.103.4.1289.
Rosenqvist, Erik. 2018. "Two Functions of Peer Influence on Upper-Secondary Education Application Behavior." *Sociology of Education* 91 (1): 72–89. https://doi.org/10.1177/0038040717746113.
Sacerdote, Bruce I. 2011. "Chapter 4 – Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know." *Handbook of the Economics of Education / Ed. by Eric A. Hanushek, Stephen Machin and Ludger* 3 (11): 249–77. http://www.gtbib.net/gtbib.php?aplicacion=sod&tipo=consulta&plantilla=usuario_peticiones.php&tabla=peticiones&relaciones[]=peticiones_refbib&pestanya=&campo=PETICION&texto=13502055&operadores[sod_peticiones.CODI_USU]=IGU&campos[sod_peticiones.CODI_USU]=!G.
———. 2014. "Experimental and Quasi-Experimental Analysis of Peer Effects: Two Steps Forward?" *Annual Review of Economics* 6 (1): 253–72. https://doi.org/10.1146/annurev-economics-071813-104217.
Seaton, Marjorie, Herbert W. Marsh, and Rhonda G. Craven. 2009. "Earning Its Place as a Pan-Human Theory: Universality of the Big-Fish-Little-Pond Effect Across 41 Culturally and Economically Diverse Countries." *Journal of Educational Psychology* 101 (2): 403–19. https://doi.org/10.1037/a0013838.
Seymour, Elaine. 1995. "The Loss of Women from Science, Mathematics, and Engineering

Undergraduate Majors: An Explanatory Account." *Science Education* 79 (4): 437–73. https://doi.org/10.1002/sce.3730790406.
Shavelson, Richard J, Judith J Hubner, and George C Stanton. 1976. "Self-Concept: Validation of Construct Interpretations." *Review of Educational Research* 46 (3): 407. https://doi.org/10.2307/1170010.
Stouffer, Samuel A, Edward A Suchman, Leland C DeVinney, Shirley A Star, and Robin M Williams Jr. 1949. "The American Soldier: Adjustment during Army Life.(Studies in Social Psychology in World War Ii), Vol. 1."
Suls, Jerry. 2000. *Handbook of Social Comparison Theory and Research*. Edited by Jerry Suls and Ladd Wheeler. New York: Springer Science+Business Media. http://journals.sagepub.com/doi/10.1177/001872675400700202.
Suls, Jerry, René Martin, and Ladd Wheeler. 2002. "Social Comparison: Why, with Whom, and with What Effect?" *Current Directions in Psychological Science* 11 (5): 159–63. https://doi.org/10.1111/1467-8721.00191.
Trautwein, Ulrich, Oliver Lüdtke, Brent W. Roberts, Inge Schnyder, and Alois Niggli. 2009. "Different Forces, Same Consequence: Conscientiousness and Competence Beliefs Are Independent Predictors of Academic Effort and Achievement." *Journal of Personality and Social Psychology* 97 (6): 1115–28. https://doi.org/10.1037/a0017048.
Veblen, Thorstein. 1899. *The Theory of the Leisure Class*. New York: Macmillan.
Vinni-Laakso, Janica, Jiesi Guo, Kalle Juuti, Anni Loukomies, Jari Lavonen, and Katariina Salmela-Aro. 2019. "The Relations of Science Task Values, Self-Concept of Ability, and Stem Aspirations among Finnish Students from First to Second Grade." *Frontiers in Psychology* 10 (JUL): 1–15. https://doi.org/10.3389/fpsyg.2019.01449.
Wayment, Heidi A., and Shelley E. Taylor. 1995. "Self-Evaluation Processes: Motives, Information Use, and Self-Esteem." *Journal of Personality* 63 (4): 729–57. https://doi.org/10.1111/j.1467-6494.1995.tb00315.x.
Wills, Thomas A. 1981. "Downward Comparison Principles in Social Psychology." *Psychological Bulletin* 90 (2): 245–71. https://doi.org/10.1037/0033-2909.90.2.245.

Appendix A: Additional Tables and Figures

Appendix Tables

Table A1: Balance checks

	Girl	Age	Poor	Rich	Behavior	Diligence	Baseline GPA
Desk mate's baseline	0.01 (0.01)	0.02+ (0.01)	-0.00 (0.00)	0.00 (0.01)	-0.01 (0.01)	-0.00 (0.01)	0.01 (0.01)
Leave-one-out class mean	-12.68** (0.68)	-12.50** (0.67)	-12.16** (0.62)	-11.97** (0.77)	-12.58** (0.72)	-12.76** (0.69)	-12.42** (0.59)
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,014	2,967	2,865	2,865	2,846	2,847	2,908

Notes: The first line shows the OLS regression coefficient of own baseline characteristic on the deskmate's baseline characteristic. All models contain classroom fixed effects. Observations vary across models because of item-specific missing values. Standard errors are clustered at the school level (in parentheses).
** p<0.01, * p<0.05, + p<0.1

Table A2: Pre-registered estimates for the causal effects of deskmate baseline GPA on students' average academic self-concept showing full regression results for Table 2.

	(1) AASC	(2) CASC	(3) AASC	(4) CASC
DM Lower	0.03	0.05	0.03	0.06
	(0.04)	(0.05)	(0.04)	(0.05)
DM Higher	-0.04	0.00	-0.06	-0.00
	(0.06)	(0.05)	(0.06)	(0.06)
Own GPA	0.90**	0.95**	0.89**	0.94**
	(0.05)	(0.06)	(0.05)	(0.06)
Girl	-0.13*	-0.08	-0.15**	-0.11+
	(0.05)	(0.05)	(0.05)	(0.05)
Age	-0.06	-0.03	-0.05	-0.03
	(0.03)	(0.03)	(0.04)	(0.03)
Poor	-0.11+	-0.12+	-0.10+	-0.12+
	(0.05)	(0.07)	(0.05)	(0.07)
Rich	0.11+	0.10	0.11+	0.09
	(0.06)	(0.06)	(0.06)	(0.06)
Behavior missing	0.80**	0.25*	-0.21*	-0.43*
	(0.07)	(0.11)	(0.09)	(0.16)
Behavior =2	0.09	0.18	0.11	0.19
	(0.16)	(0.18)	(0.16)	(0.18)
Behavior =3	0.05	-0.00	0.07	-0.00
	(0.08)	(0.09)	(0.08)	(0.09)
Behavior =4	-0.00	-0.05	-0.00	-0.06
	(0.04)	(0.04)	(0.04)	(0.04)
Behavior = 5	Ref.	Ref.	Ref.	Ref.
Diligence missing	-0.98**	-0.65**	0.00	0.00
	(0.05)	(0.06)	(0.00)	(0.00)
Diligence = 2	-0.38**	-0.44*	-0.46**	-0.54**
	(0.14)	(0.17)	(0.14)	(0.17)
Diligence = 3	-0.32**	-0.32**	-0.37**	-0.37**
	(0.10)	(0.10)	(0.10)	(0.11)
Diligence = 4	-0.21**	-0.30**	-0.23**	-0.33**
	(0.06)	(0.07)	(0.06)	(0.07)
Diligence = 5	Ref.	Ref.	Ref.	Ref.
Roma			0.02	0.06
			(0.07)	(0.07)
Constant	2.26**	1.69**	2.23**	1.75**
	(0.42)	(0.42)	(0.43)	(0.42)
Observations	2,856	2,805	2,712	2,663
R-squared	0.61	0.58	0.62	0.59
F (DM Lower = DM Higher = 0	0.64	0.73	1.04	0.87
Two-sided p-value	0.53	0.37	0.17	0.30
F (DM Lower = DM Higher	1.16	0.81	1.95	1.09
Two-sided p-value	0.29	0.49	0.36	0.43

Note: AASC is the absolute academic self-concept. CASC is the comparative academic self-concept.
All models control for classroom fixed effects. Standard errors are clustered at the school level.
Robust standard errors in parentheses. ** p<0.01, * p<0.05, + p<0.1
Columns 1-2 report models without control variables, and columns 3-4 report models that control for baseline control variables.
Control variables: Girl (=1), Age, Poor (=1), Rich (=1), Baseline behavior grade in dummies (grade 5 is the reference category), Baseline diligence grade in dummies (grade 5 is the reference category). We code missing values in the covariates as zero and enter dummy variables to control for missingness. Missingness in the variables Rich and Poor are controlled by classroom fixed effects, as missingness in these variables affects entire classrooms.
No deskmate coefficient remains statistically significant after Benjamini-Hochberg correction for multiple testing.

Table A3: Causal effects of deskmate's GPA on students' average ASC by gender (pre-registered, exploratory analysis)

		Without controls		With controls	
		(1) AASC	(2) CASC	(3) AASC	(4) CASC
Panel A: Boy	DM Lower	-0.08 (0.07)	-0.02 (0.08)	-0.09 (0.07)	-0.03 (0.08)
	DM Higher	-0.05 (0.08)	-0.03 (0.07)	-0.04 (0.09)	-0.02 (0.07)
	Own GPA	1.06** (0.04)	1.10** (0.05)	0.95** (0.05)	0.98** (0.07)
	Constant Controls	Yes No	Yes No	Yes Yes	Yes Yes
	Observations R-squared	1,480 0.64	1,456 0.60	1,480 0.65	1,456 0.60
	F(DMLow=DMHigh=0) Two-sided p-value	0.68 0.51	0.11 0.90	0.82 0.45	0.09 0.91
	F (DM Low = DM High) Two-sided p-value	0.15 0.71	0.03 0.87	0.30 0.59	0.01 0.94
Panel B: Girl	DM Lower	0.14* (0.05)	0.15* (0.07)	0.14* (0.05)	0.14+ (0.07)
	DM Higher	-0.05 (0.09)	-0.01 (0.09)	-0.05 (0.08)	0.01 (0.08)
	Own GPA	1.00** (0.04)	1.07** (0.05)	0.85** (0.07)	0.93** (0.09)
	Constant Controls	Yes No	Yes No	Yes Yes	Yes Yes
	Observations R-squared	1,376 0.64	1,349 0.63	1,376 0.65	1,349 0.65
	F(DMLow=DMHigh=0) Two-sided p-value	5.75 0.01	2.66 0.08	5.24 0.02	2.14 0.12
	F (DM Low = DM High) Two-sided p-value	6.08 0.02	3.31 0.08	6.08 0.01	2.50 0.13

Note: Each panel is estimated separately.
AASC is the absolute academic self-concept. CASC is the comparative academic self-concept.
All models control for classroom fixed effects. Standard errors are clustered at the school level.
Robust standard errors in parentheses. ** p<0.01, * p<0.05, + p<0.1
Columns 1-2 report models without control variables, and columns 3-4 report models that control for baseline control variables.
Control variables: Girl (=1), Age, Poor (=1), Rich (=1), Baseline behavior grade in dummies (grade 5 is the reference category), Baseline diligence grade in dummies (grade 5 is the reference category).
We code missing values in the covariates as zero and enter dummy variables to control for missingness. Missingness in the variables Rich and Poor are controlled by classroom fixed effects, as missingness in these variables affects entire classrooms.
No deskmate coefficient remains statistically significant after Benjamini-Hochberg correction for multiple testing.

Table A4: Causal effects of deskmate's subject-specific baseline grades on students' subject-specific ASC by gender

		Without controls		With controls	
		(1) AASC	(2) CASC	(3) AASC	(4) CASC
Grammar					
Panel A: Boy	DM Lower in grammar	-0.00 (0.09)	0.08 (0.09)	-0.00 (0.08)	0.08 (0.10)
	DM Higher in grammar	0.01 (0.10)	0.08 (0.09)	0.04 (0.10)	0.12 (0.09)
	Own Grammar grade	0.85** (0.05)	0.91** (0.06)	0.61** (0.07)	0.61** (0.08)
	Constant Controls	Yes No	Yes No	Yes Yes	Yes Yes
	Observations R-squared	1,411 0.47	1,354 0.46	1,411 0.50	1,354 0.50
Panel B: Girl	DM Lower in grammar	0.24** (0.08)	0.15 (0.12)	0.24** (0.07)	0.14 (0.11)
	DM Higher in grammar	0.04 (0.07)	-0.03 (0.08)	0.05 (0.07)	-0.00 (0.08)
	Own Grammar grade	0.81** (0.06)	0.90** (0.07)	0.48** (0.09)	0.53** (0.10)
	Constant Controls	Yes No	Yes No	Yes Yes	Yes Yes
	Observations R-squared	1,312 0.51	1,256 0.51	1,312 0.54	1,256 0.56
Literature					
Panel A: Boy	DM Lower in Literature	-0.08 (0.11)	-0.09 (0.11)	-0.07 (0.11)	-0.08 (0.11)
	DM Higher in Literature	-0.06 (0.12)	-0.12 (0.13)	-0.05 (0.11)	-0.11 (0.12)
	Own Literature grade	0.72** (0.05)	0.78** (0.06)	0.40** (0.06)	0.48** (0.08)
	Constant Controls	Yes No	Yes No	Yes Yes	Yes Yes
	Observations R-squared	1,422 0.44	1,367 0.45	1,422 0.48	1,367 0.48
Panel B: Girl	DM Lower in Literature	-0.03 (0.11)	0.09 (0.11)	-0.04 (0.10)	0.09 (0.11)
	DM Higher in Literature	-0.25* (0.12)	-0.14 (0.12)	-0.23+ (0.13)	-0.11 (0.13)
	Own Literature grade	0.70** (0.04)	0.78** (0.05)	0.38** (0.06)	0.48** (0.09)
	Constant Controls	Yes No	Yes No	Yes Yes	Yes Yes
	Observations R-squared	1,327 0.44	1,267 0.48	1,327 0.48	1,267 0.51
Mathematics					
Panel A: Boy	DM Lower in Mathematics	0.04 (0.13)	0.07 (0.11)	0.05 (0.12)	0.08 (0.11)
	DM Higher in Mathematics	0.09 (0.13)	0.15 (0.11)	0.12 (0.13)	0.20+ (0.11)
	Own Mathematic grade	1.04** (0.07)	1.09** (0.07)	0.78** (0.08)	0.80** (0.09)
	Constant Controls	Yes No	Yes No	Yes Yes	Yes Yes

		(1)	(2)	(3)	(4)
	Observations	1,422	1,356	1,422	1,356
	R-squared	0.46	0.47	0.48	0.50
Panel B: Girl	DM Lower in Mathematics	-0.04	-0.08	-0.07	-0.11
		(0.11)	(0.11)	(0.11)	(0.11)
	DM Higher in Mathematics	-0.19	-0.11	-0.18	-0.09
		(0.14)	(0.15)	(0.14)	(0.14)
	Own Mathematic grade	0.94**	0.99**	0.63**	0.64**
		(0.07)	(0.07)	(0.11)	(0.11)
	Constant	Yes	Yes	Yes	Yes
	Controls	No	No	Yes	Yes
	Observations	1,323	1,276	1,323	1,276
	R-squared	0.47	0.46	0.50	0.50

Note: Each panel is estimated separately.
AASC is the absolute academic self-concept. CASC is the comparative academic self-concept.
All models control for classroom fixed effects. Standard errors are clustered at the school level.
Robust standard errors in parentheses. ** p<0.01, * p<0.05, + p<0.1
Columns 1-2 report models without control variables, and columns 3-4 report models that control for baseline control variables.
Control variables: Girl (=1), Age, Poor (=1), Rich (=1), Baseline behavior grade in dummies (grade 5 is the reference category), Baseline diligence grade in dummies (grade 5 is the reference category). Missingness in the covariates are coded as zero, and a dummy variable controls for missing status.
We code missing values in the covariates as zero and enter dummy variables to control for missingness.
Missingness in the variables Rich and Poor are controlled by classroom fixed effects, as missingness in these variables affects entire classrooms.
No deskmate coefficient remains statistically significant after Benjamini-Hochberg correction for multiple testing.

Table A5: The causal effects of deskmate baseline GPA on students' average academic self-concepts, by grade-level

Panel A: AASC						
	(1) Grade 3	(2) Grade 4	(3) Grade 5	(4) Grade 6	(5) Grade 7	(6) Grade 8
DM Lower	0.07	0.04	0.07	0.04	0.02	-0.11
	(0.11)	(0.12)	(0.10)	(0.09)	(0.11)	(0.10)
DM Higher	-0.02	-0.05	0.01	0.02	-0.16	-0.12
	(0.12)	(0.11)	(0.14)	(0.13)	(0.15)	(0.13)
Own GPA	1.10**	0.90**	1.04**	1.02**	0.99**	1.06**
	(0.08)	(0.07)	(0.09)	(0.06)	(0.07)	(0.08)
Constant	0.55+	1.40**	0.64+	0.88**	1.15**	0.97**
	(0.32)	(0.28)	(0.36)	(0.25)	(0.24)	(0.29)
Observations	493	522	575	456	420	390
R-squared	0.53	0.49	0.59	0.61	0.67	0.68
Panel B: CASC						
	(1) Grade 3	(2) Grade 4	(3) Grade 5	(4) Grade 6	(5) Grade 7	(6) Grade 8
DM Lower	0.10	0.03	0.04	0.12	0.02	0.02
	(0.16)	(0.13)	(0.11)	(0.11)	(0.16)	(0.12)
DM Higher	-0.00	-0.04	0.00	0.08	-0.15	-0.01
	(0.12)	(0.11)	(0.14)	(0.11)	(0.16)	(0.10)
Own GPA	1.06**	0.95**	1.07**	1.13**	1.10**	1.17**
	(0.09)	(0.06)	(0.11)	(0.09)	(0.07)	(0.07)
Constant	0.55	1.05**	0.40	0.32	0.75**	0.43+
	(0.37)	(0.23)	(0.45)	(0.31)	(0.26)	(0.24)
Observations	484	516	561	452	410	382
R-squared	0.47	0.47	0.55	0.61	0.67	0.64

Note: Each panel is estimated separately.
AASC is the absolute academic self-concept. CASC is the comparative academic self-concept.
All models control for classroom fixed effects. Standard errors are clustered at the school level.
Robust standard errors in parentheses. ** p<0.01, * p<0.05, + p<0.1
Columns 1-2 report models without control variables, and columns 3-4 report models that control for baseline control variables.
Control variables: Girl (=1), Age, Poor (=1), Rich (=1), Baseline behavior grade in dummies (grade 5 is the reference category), Baseline diligence grade in dummies (grade 5 is the reference category).
We code missing values in the covariates as zero and enter dummy variables to control for missingness. Missingness in the variables Rich and Poor are controlled by classroom fixed effects, as missingness in these variables affects entire classrooms.

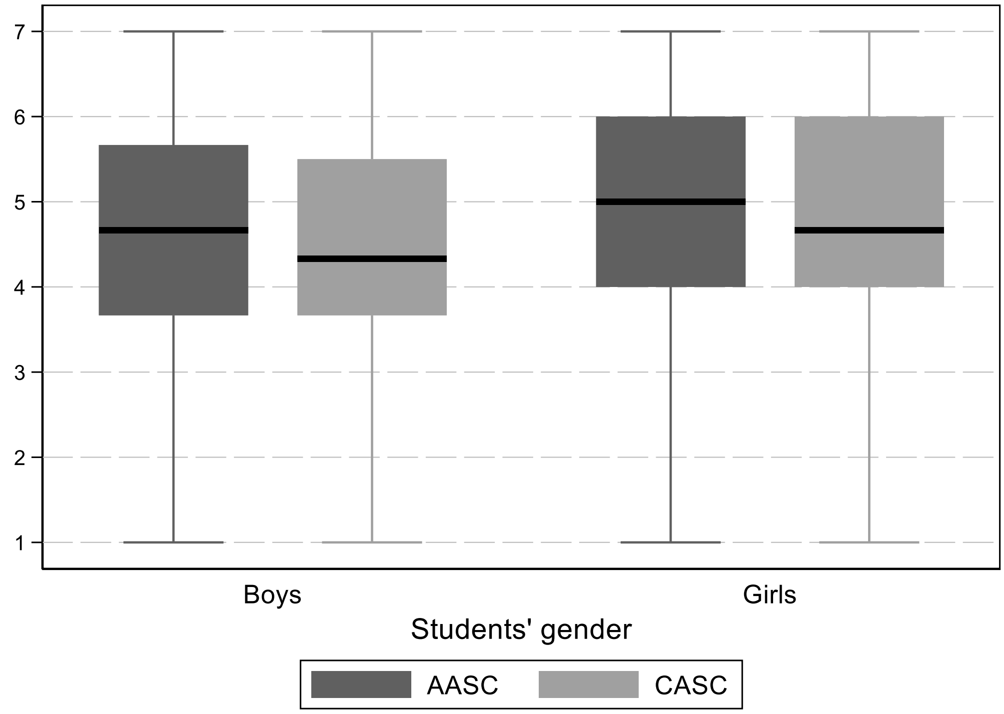
Table A6: Pre-registered interaction analysis between deskmate GPA and classroom average baseline GPA

	(1) AASC	(2) CASC	(3) AASC	(4) CASC
DM Lower	0.04	0.07	0.04	0.06
	(0.04)	(0.04)	(0.04)	(0.05)
DM Higher	-0.02	0.01	-0.02	0.01
	0.04	0.07	0.04	0.06
Own GPA	1.03**	1.09**	0.92**	0.96**
	(0.03)	(0.04)	(0.04)	(0.06)
DM Lower × classroom-av. GPA	0.04	-0.00	0.04	-0.00
	(0.08)	(0.10)	(0.08)	(0.10)
DM Higher × classroom-av. GPA	0.09	-0.02	0.11	0.01
	(0.10)	(0.08)	(0.11)	(0.09)
Own GPA × classroom-av. GPA	0.14*	0.09	0.14*	0.06
	(0.06)	(0.08)	(0.07)	(0.09)
Constant	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes
Observations	2,856	2,805	2,856	2,805
R-squared	0.61	0.58	0.61	0.58
F (DM Lower = DM Higher	1.13	1.35	0.74	0.72
Two-sided p-value	0.29	0.25	0.39	0.40

AASC is the absolute academic self-concept. CASC is the comparative academic self-concept.
All models control for classroom fixed effects. Standard errors are clustered at the school level.
Robust standard errors in parentheses. ** p<0.01, * p<0.05, + p<0.1
Columns 1-2 report models without control variables, and columns 3-4 report models that control for baseline control variables.
Control variables: Girl (=1), Age, Poor (=1), Rich (=1), Baseline behavior grade in dummies (grade 5 is the reference category), Baseline diligence grade in dummies (grade 5 is the reference category).
We code missing values in the covariates as zero and enter dummy variables to control for missingness. Missingness in the variables Rich and Poor are controlled by classroom fixed effects, as missingness in these variables affects entire classrooms.

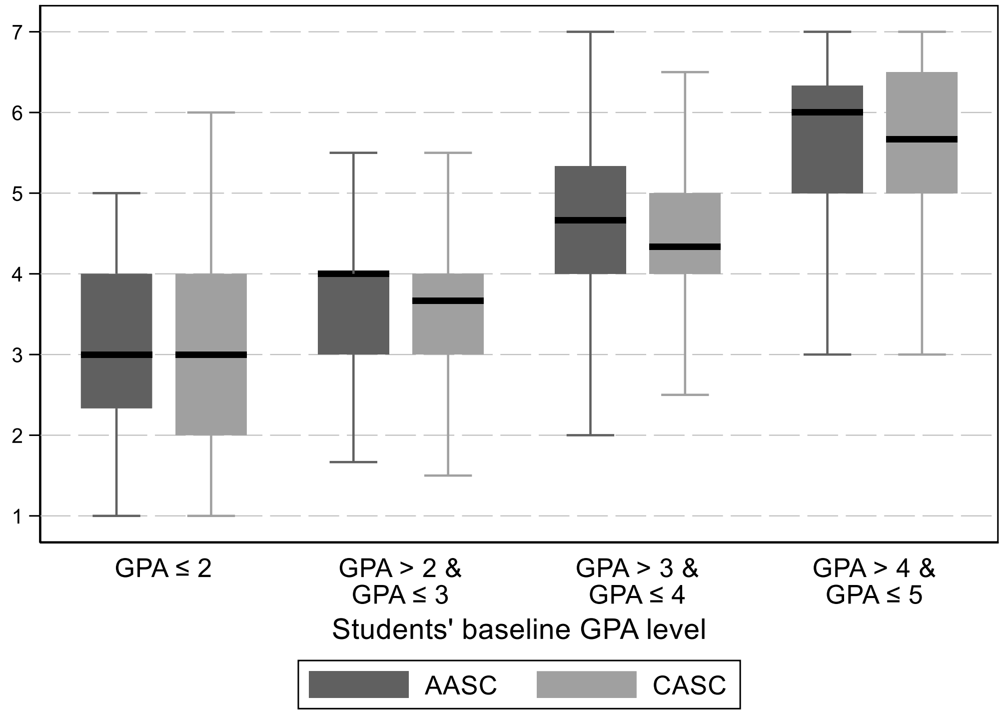
Appendix Figures

Figure A1: Outcome distributions by students' gender (average AASC and CASC)



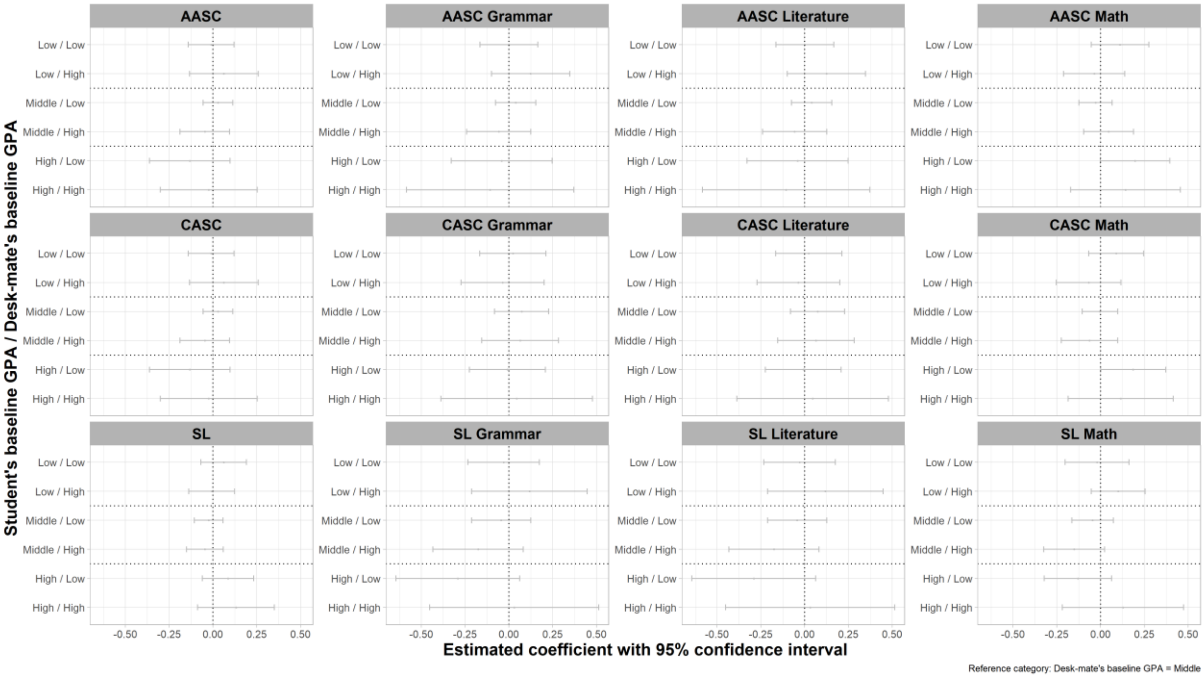
Note: Boxes represent data between the 1st and 3rd quartiles; center lines represent medians; leaves give minima and maxima.

Figure A2: Outcome distributions by students' baseline GPA levels (average AASC and CASC)



Note: Boxes represent data between the 1st and 3rd quartiles; center lines represent medians; leaves give minima and maxima. GPA refers to students' own baseline GPA.

Figure A3: Exploring heterogeneity in the effect of deskmates' relative GPA according to students' own GPA (equation 2) by subject.



Note: Each subgraph gives conventional (unpenalized) 95% confidence intervals for the estimated effects of sitting next to a deskmate with low or high grades (compared to a deskmate with middle grades) among students who themselves have low, middle, or high grades within their classroom, respectively. No deskmate coefficient remains statistically significant after Benjamini-Hochberg correction for multiple testing.

Appendix B: Characteristics of participating schools compared to the national average

We characterize the representativeness of our analytic sample by comparing students in participating schools to students nationwide using data from the 2017 National Assessment of Basic Competencies (NABC). The NABC database contains two types of data. First, reading comprehension and mathematics test scores for all students, assessed by a mandatory, supervised, and standardized PISA-like, in-school test (mean=0, SD=1). Second, students' social background information collected via a voluntary take-home survey. We focus the comparison on students who were in 6th grade in May 2017, a few months before baseline.

Table B1 shows that included students in included schools had lower standardized test scores in math and reading, and students' parents were less educated than the national average. The parents of students in included schools were about as likely to be employed as students' parents nationally.

Table B1.: Student test scores and parental socio-economic status of 6th-grade students in included schools and nationwide

	Math test	Reading test	Mother has a college degree	Mother works	Father has a college degree	Father works
Students in participating schools	-0.19	-0.26	0.18	0.73	0.12	0.86
All students	0.00	0.00	0.31	0.78	0.24	0.87

Data: NABC (National Assessment of Basic Competences), 6th grade, May 2017.

Appendix C: Results for the outcome 'subject liking'

Table C1.: Pre-registered estimates for the causal effects of deskmate baseline GPA on students' subject liking

	Without controls	With controls
	(1)	(2)
DM Lower	-0.02	-0.03
	(0.04)	(0.04)
	[-0.03]	[-0.03]
DM Higher	-0.07+	-0.08+
	(0.04)	(0.04)
	[-0.09]	[-0.10]
Own GPA	0.26**	0.18**
	(0.02)	(0.03)
Constant	Yes	Yes
Mean of the dependent variable	3.590	3.590
Observations	2,884	2,884
R-squared	0.34	0.35
F (DM Lower = DM Higher = 0	1.50	1.64
Two-sided p-value	0.26	0.21
F (DM Lower = DM Higher	1.31	1.25
Two-sided p-value	0.24	0.27

AASC is the absolute academic self-concept. CASC is the comparative academic self-concept.
All models control for classroom fixed effects. Standard errors are clustered at the school level.
Robust standard errors in parentheses. ** p<0.01, * p<0.05, + p<0.1
Cohen's D effect sizes (the coefficient divided by the standard deviation) are in [sqaured breackets]
Columns 1-2 report models without control variables, and columns 3-4 report models that control for baseline control variables.
Control variables: Girl (=1), Age, Poor (=1), Rich (=1), Baseline behavior grade in dummies (grade 5 is the reference category), Baseline diligence grade in dummies (grade 5 is the reference category).
We code missing values in the covariates as zero and enter dummy variables to control for missingness. Missingness in the variables Rich and Poor are controlled by classroom fixed effects, as missingness in these variables affects entire classrooms.
No deskmate coefficient remains statistically significant after Benjamini-Hochberg correction for multiple testing.

Table C2: Estimated causal effects of deskmates' subject-specific baseline grades on students' subject liking in the same subject, by school subject (pre-registered exploratory analyses)

		(1) Without controls	(2) With controls
Panel A: Grammar	DM Lower in grammar	0.00 (0.05) [0.00]	0.00 (0.05) [0.00]
	DM Higher in grammar	0.01 (0.07) [0.01]	0.01 (0.06) [0.01]
	Own Grammar grade	0.24** (0.03)	0.15** (0.03)
	Constant	Yes	Yes
	Mean of the dependent variable	3.47	3.47
	Controls	No	No
	Observations	2,843	2,843
	R-squared	0.45	0.44
	F(DMLow=DMHigh=0)	0.01	0.01
	Two-sided p-value	0.99	0.94
	F (DM Low = DM High)	0.02	0.01
	Two-sided p-value	0.90	0.99
Panel B: Literature	DM Lower in literature	0.10* (0.04) [0.10]	0.10* (0.04) [0.09]
	DM Higher in literature	0.02 (0.05) [0.02]	0.01 (0.06) [0.01]
	Own Literature grade	0.26** (0.02)	0.19** (0.03)
	Constant	Yes	Yes
	Mean of the dependent variable	3.69	3.69
	Controls	No	No
	Observations	2,835	2,835
	R-squared	0.28	0.29
	F(DMLow=DMHigh=0)	2.66	2.54
	Two-sided p-value	0.08	0.15
	F (DM Low = DM High)	1.93	2.15
	Two-sided p-value	0.17	0.09
Panel C: Mathematics	DM Lower in mathematics	-0.07 (0.05) [-0.06]	-0.07 (0.05) [-0.06]
	DM Higher in mathematics	-0.04 (0.07) [-0.03]	-0.01 (0.07) [-0.01]
	Own Mathematics grade	0.22** (0.04)	0.12** (0.04)
	Constant	Yes	Yes
	Mean of the dependent variable	3.62	3.62
	Controls	No	No
	Observations	2,834	2,834
	R-squared	0.28	0.30
	F(DMLow=DMHigh=0)	0.89	0.90
	Two-sided p-value	0.42	0.42
	F (DM Low = DM High)	0.26	0.66
	Two-sided p-value	0.61	0.42

Note: Each panel is estimated separately.
AASC is the absolute academic self-concept. CASC is the comparative academic self-concept.
All models control for classroom fixed effects. Standard errors are clustered at the school level.
Robust standard errors in parentheses. ** p<0.01, * p<0.05, + p<0.1
Cohen's D effect sizes (the coefficient divided by the standard deviation) are in [sqaured breackets]
Columns 1-2 report models without control variables, and columns 3-4 report models that control for baseline control variables.
Control variables: Girl (=1), Age, Poor (=1), Rich (=1), Baseline behavior grade in dummies (grade 5 is the reference category), Baseline diligence grade in dummies (grade 5 is the reference category).
We code missing values in the covariates as zero and enter dummy variables to control for missingness.
Missingness in the variables Rich and Poor are controlled by classroom fixed effects, as missingness in these variables affects entire classrooms.
No deskmate coefficient remains statistically significant after Benjamini-Hochberg correction for multiple testing.

Appendix D: Alterntive measures of baseline achievement

Our pre-registered analyses measured baseline achievement by student's baseline GPA. After the conclusion of our study, we gained acceess to comprehensive, nationally standardized test scores from the National Assessment of Basic Competencies (NABC, described above). The NABC is only conducted in certain grade levels, including 6th grade. We obtained parental consent to link 71 percent of 7th graders and 90.5 percent of 8th graders in our sample to their 6th-grade NABC scores. The NABC is a state-of-the-art achievement test, similar to the PISA, and hence arguably serves as a better, less error-laden, measure of students' baseline achievement.

We used NABC test scores to compute alternative measures of baseline achievement: the average of students' 6th-grade mathematics and reading comprehension tests, standardized to mean-0, SD=1. We classified the relative performance of students and deskmates as the difference between students' and deskmates' average scores in three different ways using benchmarks of 0.15, 0.25, and 0.5 standard deviation difference, respectively. For example, using the benchmark of 0.15 SD, deskmates' NABC score is coded as "lower" if the deskmates NABC score is more than 0.15 SDs lower than students' own NABC score.

We found no effect of deskmates' relative baseline achievement on students' average AASC or CASC using any of our three alterantive definitions of relative baseline achievement difference. Point estimates are substantively small and not statistically different from zero (Table D1). To the extent that these alternative measures of baseline achievement contain less random measurement error than our primary GPA-based measure of baseline achievement, this indicates that the Null results reported in the main body of the paper are not due to atenuation due to measurement error.

Table D1: The causal effects of deskmate's 6th-grade baseline NABC score on students' average academic self-concept – an analysis of the sample of 7th and 8th-grade students [not pre-registered analysis]

	0.15 SD benchmark		0.25 SD benchmark		0.50 SD benchmark	
	(1) AASC	(2) CASC	(3) AASC	(4) CASC	(5) AASC	(6) CASC
Own NABC score	0.91** (0.08)	1.03** (0.08)	0.92** (0.08)	1.04** (0.08)	0.93** (0.08)	1.05** (0.08)
DM's NABC lower	0.16 (0.20)	0.05 (0.20)	0.17 (0.17)	0.02 (0.17)	0.01 (0.15)	-0.08 (0.15)
DM's NABC higher	0.07 (0.17)	-0.01 (0.18)	0.11 (0.14)	-0.02 (0.15)	-0.03 (0.12)	-0.10 (0.12)
Constant	4.33** (0.15)	4.37** (0.16)	4.32** (0.11)	4.39** (0.12)	4.44** (0.07)	4.45** (0.07)
Observations	612	599	612	599	612	599
R-squared	0.48	0.46	0.48	0.46	0.47	0.46

Note: All models control for classroom fixed effects, but no further control variables. Standard errors are clustered at the school level. Robust standard errors in parentheses. ** p<0.01, * p<0.05, + p<0.1.

Appendix E: Supplementary student and teacher surveys

We conducted three supplementary teacher surveys and one student survey to learn about the design of seating charts, deskmate interactions, and student-to-student comparisons in Hungarian primary schools between 2017 and 2022. These surveys were not pre-registered and were conducted either before or after our field experiment.

1. Teacher survey, 2017

In order to learn about the typical design of seating charts in participating classrooms, we conducted an online survey among the primary-school homeroom teachers who would subsequently participate in the field experiment. The survey was conducted between 22 May and 18 July 2017. We collected data from 161 teachers (out of the 195 who would participate in the experiment) in 41 schools.

Responding teachers were mostly female (74%), 48.8 years old (SD = 9.6), and had 22.7 years of teaching experience (SD = 10.3) on average. Most taught Hungarian (53%) or mathematics (46%).

Most teachers reported designing the teaching chart themselves. Specifically, 66% of respondents indicated that “As a homeroom teacher, I decide where each student should sit”, and a further 29% said that “Some students can sit where they like, some, however, I instruct as to where they should sit”.

2. Teacher survey, 2021

In order to learn about teacher’s perceptions of deskmate relationships and seating chart partices, we conducted a nationwide online survey among primary-school homeroom teachers, focussing on teachers in grades 5-8. We sent invitation e-mails to 2,743 primary schools. 413 teachers from 266 schools participated.

Responding teachers were mostly female (89%), 49.9 years old (SD = 8.9) and had 21.9 years of teaching experience on average(SD = 10.5), and mainly taught Hungarian (47%) or mathematics (53%).

Participating schools were nationally representative with respect to: share of poor, rich, and Roma students, average math and reading comprehension scores and students’ SES-index of 8th graders, and type of settlement (assessed against administrative data from the National Assessment of Basic Competences, 8th grade sample, 2021).

Results indicate that teachers believe in deskmate effects. For example, teachers believe that deskmates affect each other’s behavior (80%), and diligence (57%) and influence each other academically (52%).

Table E1. Teachers’ perceptions of deskmate influence

Dimension of influence	academic achievement	friendship formation	school behavior	diligence	academic self-concept
Mean value	52.54%	37.29%	80.15%	57.14%	33.41%

Note: The question “To what extent do deskmates affect [dimension]?” was scored from “1 = Not at all” to “5 = In a great extent”. We coded 4 & 5 as a positive belief in influence (=1) and the other categories (1, 2, and 3) as disagreement (=0).

Seating charts are stable in Hungarian schools. 74 percent of responding teachers employed a seating chart. The majority (52%) of teachers who employ a seating chart said that they changed the seating chart only once each semester. Only 11% of teachers changed the seating chart at least every month.

3. Teacher survey, 2022

In order to learn about collaboration between deskmates in Hungarian primary schools we conducted a nationwide online survey among primary-school homeroom teachers teaching grades 1-8. We conducted this survey between 10 and 28 February 2022. We sent invitation e-mails to 1,892 primary schools and received answers from 656 teachers in 288 schools.

Respondent teachers were mostly female (92%), 49.6 years old (SD = 8.6) on average, and mainly taught Hungarian (53%) or mathematics (46%).

Participating schools were nationally representative with respect to: share of poor, rich, and Roma students, average math and reading comprehension scores, and students' SES-index of 8th graders. (assessed against administrative data from the National Assessment of Basic Competences, 8th grade sample, 2021). Village schools were somewhat overrepresented in the sample (47 vs 40 percent).

As Figure E1 shows, deskmates collaborate regularly.

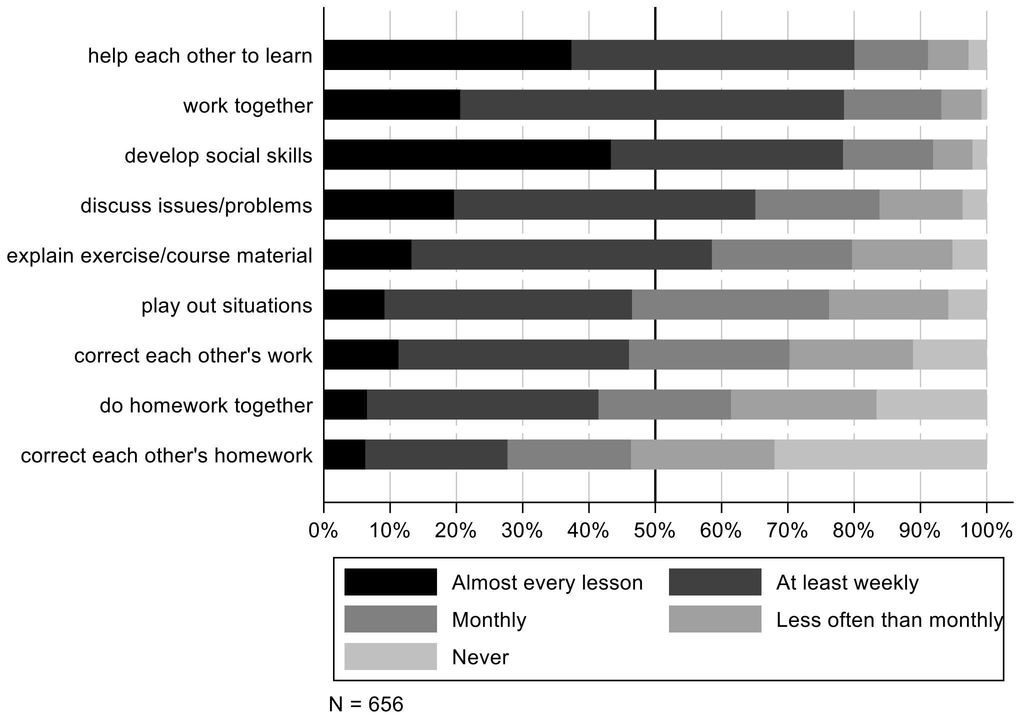


Figure E1, Deskmate collaboration

4. Student survey, 2020

In norder to learn learn to whom students compared themselves academically, we conducted an online survey among Hungarian primary school studens in grades 4-7 between 17 December 2020 and 19 February 2021. 473 students from 51 classrooms in 15 schools responded.

Respondent were 47% female, 12 years old (SD = 1.1) on average, and had an average GPA of 3.6 (SD = 0.95) on a scale ranging between 1 and 5, where 1 is the worst and 5 is the best grade.

Table E2. Percent of students who compare their own achievement to that of specific peers, shown in each column of the table

	Best student	Worst studnet	Avearge student	Friends	Deskmates
Percentage of students who compare themselves to	56.32%	25.26%	45.79%	87.37%	72.11%

Note: These questions were asked from students who compared their achievement to their peers (40% of students in the sample).