

Supplement to:

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

Tamás Keller – Felix Elwert

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

Online Supplement

Version: September 29, 2023

Appendix A: Supplementary Tables (Pre-registered)

Table A1. Covariate balance for students' and their deskmates' baseline characteristics in the pre-registered samples of the two primary outcomes

VARIABLES	(1) Female	(2) Roma	(3) Poor	(4) Rich	(5) Age	(6) Grammar	(7) Literature	(8) Mathematics	(9) Baseline GPA	(10) Behavior	(11) Diligence
PANEL A: READING SCORE											
Deskmate's baseline	0.00 (0.01)	0.00 (0.01)	-0.00 (0.00)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.00 (0.01)	0.02* (0.01)	0.02 (0.01)	0.00 (0.01)	0.00 (0.01)
Leave-one-out mean	-14.68† (0.59)	-13.50† (0.83)	-13.76† (0.63)	-13.97† (0.57)	-14.08† (0.69)	-13.83† (0.61)	-14.05† (0.60)	-14.01† (0.57)	-13.90† (0.62)	-13.59† (0.62)	-13.66† (0.54)
Classroom fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,909	2,762	2,762	2,762	2,870	2,887	2,889	2,881	2,909	2,891	2,893
R-squared	0.90	0.90	0.90	0.91	0.90	0.89	0.90	0.90	0.90	0.89	0.91
PANEL B: ENDLINE GPA											
Deskmate's baseline	0.00 (0.01)	0.00 (0.01)	-0.00 (0.00)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.00 (0.01)	0.02* (0.01)	0.01 (0.01)	0.00 (0.01)	0.00 (0.01)
Leave-one-out mean	-14.66† (0.59)	-13.51† (0.80)	-13.73† (0.59)	-13.83† (0.56)	-14.09† (0.65)	-13.70† (0.57)	-13.93† (0.58)	-13.96† (0.54)	-13.79† (0.58)	-13.55† (0.60)	-13.66† (0.53)
Classroom fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,108	2,964	2,964	2,964	3,034	3,084	3,086	3,079	3,108	3,090	3,092
R-squared	0.90	0.90	0.90	0.91	0.90	0.89	0.90	0.90	0.90	0.89	0.91

Note: All models control for classroom fixed effects and include a constant. No additional controls are included. Robust standard errors clustered at the school level in parentheses. † p<0.01, * p<0.05.

Table A2. Deskmate effects on students' achievement by deskmate's baseline GPA, with additional baseline covariates

	Primary outcomes		Secondary outcomes					
	(1) Reading score	(2) Endline GPA	(3) Endline Grammar	(4) Endline Literature	(5) Endline Mathematics	(6) Endline Behavior	(7) Endline Diligence	(8) Scores on teacher-written tests
Own baseline GPA	8.76 [†] (0.77)	0.72 [†] (0.02)	0.72 [†] (0.03)	0.64 [†] (0.03)	0.81 [†] (0.03)	0.06* (0.02)	0.40 [†] (0.02)	14.19 [†] (0.54)
Deskmate's baseline GPA	0.30 (0.36)	0.02 (0.01)	0.03* (0.02)	0.00 (0.01)	0.03 (0.02)	0.02 (0.02)	0.04 ^{†§} (0.01)	0.69* (0.28)
Female deskmate	2.89 [†] (0.71)	0.01 (0.02)	0.15 [†] (0.03)	0.10 [†] (0.02)	-0.21 [†] (0.03)	0.17 [†] (0.02)	0.03 (0.02)	-0.65 (0.48)
Classroom fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of the outcome	54.91	3.58	3.56	3.72	3.47	4.20	3.88	67.31
Standardized deskmate effect	0.01	0.02	0.03	0.00	0.02	0.03	0.04	0.03
Observations	2,909	3,108	3,096	3,068	3,076	3,110	3,113	2,686
R-squared	0.23	0.76	0.64	0.63	0.63	0.57	0.65	0.71

Note: Pre-registered robustness checks. This table re-estimates the models of Table 2 with additional baseline controls (students' baseline age, ethnicity, and teacher-reported socioeconomic status). Missing values in the covariates were coded as zero, and we added missingness dummies. GPA is computed from grades in grammar, literature, and mathematics (5 is best, 1 is worst). All models control for classroom fixed effects and include a constant. Standardized deskmate effects equal the deskmate coefficient divided by the standard deviation of the outcome. Robust standard errors are clustered at the school level in parentheses. † p<0.01, * p<0.05.

§The deskmate coefficient for the outcome Endline Diligence remains statistically significant at a 5% false-discovery rate after correcting for multiple hypothesis testing using Benjamini and Hochberg's (1995) procedure, treating the deskmate coefficients in columns 3-9 as one set. No coefficients remain significant when considering all deskmate coefficients for secondary outcomes in this table and in Table 2 as one set of tests.

Table A3. Deskmate effects on students' achievement by deskmate's gender, with additional baseline covariates

	Primary outcomes		Secondary outcomes					(8) Scores on teacher-written tests
	(1) Reading score	(2) Endline GPA	(3) Endline Grammar	(4) Endline Literature	(5) Endline Mathematics	(6) Endline Behavior	(7) Endline Diligence	
Female student	2.91† (0.69)	0.02 (0.02)	0.15† (0.03)	0.11† (0.02)	-0.20† (0.03)	0.17† (0.02)	0.03 (0.02)	-0.60 (0.44)
Female deskmate	0.32 (0.56)	0.04 (0.02)	0.06* (0.02)	0.04 (0.02)	0.02 (0.02)	0.01 (0.02)	0.02 (0.01)	0.62* (0.29)
Own baseline GPA	8.54† (0.83)	0.72† (0.02)	0.73† (0.03)	0.64† (0.03)	0.80† (0.03)	0.06† (0.02)	0.39† (0.02)	14.16† (0.54)
Classroom fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of the outcome for those with a male deskmate	54.20	3.54	3.52	3.68	3.43	4.19	3.85	66.59
Standardized deskmate effect	0.02	0.03	0.05	0.04	0.02	0.01	0.02	0.03
Observations	2,978	3,182	3,169	3,136	3,150	3,185	3,188	2,756
R-squared	0.23	0.76	0.64	0.63	0.63	0.57	0.65	0.71

Note: Pre-registered robustness checks. This table re-estimates the models of Table 3 with additional baseline controls (students' baseline age, ethnicity, teacher-reported socioeconomic status). Missing values in the covariates were coded as zero, and we added missingness dummies. GPA is computed from grades in grammar, literature, and mathematics (5 is best, 1 is worst). All models control for classroom fixed effects and include a constant. Standardized deskmate effects equal the deskmate coefficient divided by the standard deviation of the outcome. Robust standard errors are clustered at the school level in parentheses. † p<0.01, * p<0.05. No deskmate coefficient remains statistically significant at a 5% false-discovery rate after correcting for multiple hypothesis testing using Benjamini and Hochberg's (1995) procedure, treating the deskmate coefficients in columns 3-9 as one set of tests.

Table A4: Exploring alternative treatment definitions—regressing students’ subject-specific grades on deskmates’ corresponding baseline grades

	(1)	(2)	(3)	(4)	(5)
	Endline Grammar	Endline Literature	Endline Mathematics	Endline Diligence	Endline Behavior
Corresponding own baseline	0.74† (0.03)	0.65† (0.02)	0.78† (0.02)	0.78† (0.02)	0.74† (0.02)
Corresponding DM’s baseline	0.03* (0.02)	0.01 (0.01)	0.02 (0.02)	0.03* (0.01)	0.01 (0.01)
Female deskmate	0.14† (0.03)	0.27† (0.02)	-0.15† (0.03)	0.07* (0.02)	0.15† (0.02)
Classroom fixed effects	Yes	Yes	Yes	Yes	Yes
Mean of the outcome	3.56	3.72	3.47	3.88	4.20
Standardized DM effect	0.03	0.01	0.02	0.03	0.01
Observations	3,072	3,046	3,055	3,110	3,111
R-squared	0.53	0.49	0.51	0.57	0.54

Note: All models control for classroom fixed effects and include a constant. No additional controls are included. For all grades, 5 is best, and 1 is worst. Standardized deskmate effects equal the deskmate coefficient divided by the standard deviation of the outcome. Robust standard errors are clustered at the school level in parentheses. Table reports conventional (unpenalized) hypothesis tests: † p<0.01, * p<0.05. Penalizing significance levels for multiple hypothesis testing by Benjamini and Hochberg’s (1995) procedure with a 5% false-discovery rate across deskmate coefficients indicates no statistically significant coefficients.

Table A5. Multiple-hypothesis testing corrections for Figure 1

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Category	Student	DM	Dependent variable	Std. DM effect	SE	p-value	BH rank (r)	BH 5% value	(7) – (9)	BH Significant
IL	Low	Low	Reading test score	0.05	0.07	0.5106	5	0.0417	0.4689	0
IM (ref)	Low	Middle	Reading test score	0.00						
IH	Low	High	Reading test score	0.13	0.06	0.0389	1	0.0083	0.0306	0
mL	Middle	Low	Reading test score	-0.06	0.05	0.1684	3	0.0250	0.1434	0
mM (ref)	Middle	Middle	Reading test score	0.00						
mH	Middle	High	Reading test score	-0.09	0.06	0.0996	2	0.0167	0.0829	0
hL	High	Low	Reading test score	0.07	0.06	0.2285	4	0.0333	0.1951	0
hM (ref)	High	Middle	Reading test score	0.00						
hH	High	High	Reading test score	0.04	0.13	0.7412	6	0.0500	0.6912	0
IL	Low	Low	GPA	0.02	0.04	0.6546	5	0.0417	0.6130	0
IM (ref)	Low	Middle	GPA	0.00						
IH	Low	High	GPA	0.15	0.06	0.0212	2	0.0167	0.0045	0
mL	Middle	Low	GPA	-0.01	0.03	0.7202	6	0.0500	0.6702	0
mM (ref)	Middle	Middle	GPA	0.00						
mH	Middle	High	GPA	-0.06	0.05	0.2219	3	0.0250	0.1969	0
hL	High	Low	GPA	-0.21	0.06	0.0013	1	0.0083	-0.0071	1
hM (ref)	High	Middle	GPA	0.00						
hH	High	High	GPA	0.08	0.07	0.2418	4	0.0333	0.2085	0
IL	Low	Low	Grammar grade	0.07	0.05	0.2081	17	0.0236	0.1845	0
IM (ref)	Low	Middle	Grammar grade	0.00						
IH	Low	High	Grammar grade	0.20	0.07	0.0082	4	0.0056	0.0027	0
mL	Middle	Low	Grammar grade	-0.02	0.03	0.6073	25	0.0347	0.5726	0
mM (ref)	Middle	Middle	Grammar grade	0.00						
mH	Middle	High	Grammar grade	-0.07	0.05	0.1668	14	0.0194	0.1473	0
hL	High	Low	Grammar grade	-0.20	0.07	0.0065	3	0.0042	0.0023	0
hM (ref)	High	Middle	Grammar grade	0.00						
hH	High	High	Grammar grade	0.21	0.09	0.0217	7	0.0097	0.0119	0

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Category	Student	DM	Dependent variable	Std. DM effect	SE	p-value	BH rank (r)	BH 5% value	(7) – (9)	BH Significant
IL	Low	Low	Literature grade	0.02	0.05	0.6321	26	0.0361	0.5959	0
IM (ref)	Low	Middle	Literature grade	0.00						
IH	Low	High	Literature grade	0.10	0.06	0.1129	12	0.0167	0.0962	0
mL	Middle	Low	Literature grade	0.02	0.03	0.3773	21	0.0292	0.3481	0
mM (ref)	Middle	Middle	Literature grade	0.00						
mH	Middle	High	Literature grade	-0.02	0.05	0.6372	27	0.0375	0.5997	0
hL	High	Low	Literature grade	-0.20	0.06	0.0009	2	0.0028	-0.0019	1
hM (ref)	High	Middle	Literature grade	0.00						
hH	High	High	Literature grade	0.01	0.08	0.9269	35	0.0486	0.8783	0
IL	Low	Low	Mathematics grade	-0.06	0.04	0.1977	16	0.0222	0.1754	0
IM (ref)	Low	Middle	Mathematics grade	0.00						
IH	Low	High	Mathematics grade	0.10	0.07	0.1594	13	0.0181	0.1413	0
mL	Middle	Low	Mathematics grade	-0.03	0.04	0.4218	23	0.0319	0.3898	0
mM (ref)	Middle	Middle	Mathematics grade	0.00						
mH	Middle	High	Mathematics grade	-0.07	0.06	0.2683	19	0.0264	0.2420	0
hL	High	Low	Mathematics grade	-0.18	0.07	0.0106	5	0.0069	0.0037	0
hM (ref)	High	Middle	Mathematics grade	0.00						
hH	High	High	Mathematics grade	0.01	0.09	0.9113	34	0.0472	0.8641	0
IL	Low	Low	Behavior grade	0.02	0.07	0.7449	32	0.0444	0.7004	0
IM (ref)	Low	Middle	Behavior grade	0.00						
IH	Low	High	Behavior grade	0.10	0.07	0.1678	15	0.0208	0.1470	0
mL	Middle	Low	Behavior grade	-0.02	0.05	0.7271	31	0.0431	0.6840	0
mM (ref)	Middle	Middle	Behavior grade	0.00						
mH	Middle	High	Behavior grade	0.08	0.07	0.2830	20	0.0278	0.2553	0
hL	High	Low	Behavior grade	-0.03	0.06	0.6784	28	0.0389	0.6395	0
hM (ref)	High	Middle	Behavior grade	0.00						
hH	High	High	Behavior grade	0.16	0.10	0.1092	11	0.0153	0.0939	0

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Category	Student	DM	Dependent variable	Std. DM effect	SE	p-value	BH rank (r)	BH 5% value	(7) - (9)	BH Significant
IL	Low	Low	Diligence grade	-0.04	0.05	0.4072	22	0.0306	0.3766	0
IM (ref)	Low	Middle	Diligence grade	0.00						
IH	Low	High	Diligence grade	0.07	0.06	0.2675	18	0.0250	0.2425	0
mL	Middle	Low	Diligence grade	-0.06	0.03	0.0744	10	0.0139	0.0605	0
mM (ref)	Middle	Middle	Diligence grade	0.00						
mH	Middle	High	Diligence grade	-0.04	0.06	0.5787	24	0.0333	0.5453	0
hL	High	Low	Diligence grade	-0.18	0.05	0.0005	1	0.0014	-0.0009	1
hM (ref)	High	Middle	Diligence grade	0.00						
hH	High	High	Diligence grade	0.18	0.08	0.0240	8	0.0111	0.0129	0
IL	Low	Low	Teacher-written tests	-0.01	0.06	0.8027	33	0.0458	0.7569	0
IM (ref)	Low	Middle	Teacher-written tests	0.00						
IH	Low	High	Teacher-written tests	0.16	0.06	0.0182	6	0.0083	0.0099	0
mL	Middle	Low	Teacher-written tests	-0.01	0.03	0.6916	29	0.0403	0.6513	0
mM (ref)	Middle	Middle	Teacher-written tests	0.00						
mH	Middle	High	Teacher-written tests	0.00	0.06	0.9891	36	0.0500	0.9391	0
hL	High	Low	Teacher-written tests	-0.12	0.07	0.0670	9	0.0125	0.0545	0
hM (ref)	High	Middle	Teacher-written tests	0.00						
hH	High	High	Teacher-written tests	0.03	0.09	0.7090	30	0.0417	0.6673	0

Note: Grey background means „significant after Benjamini-Hochberg (1995) correction for multiple hypothesis testing, with a 5% false-discovery rate.” As in all analyses, we consider the coefficients in each primary-outcome model as its own set (i.e., 6 tests each in panels 1 and 2 of Figure 1), and the collection of deskmate tests across all secondary outcomes another set (36 tests across panels 3-8 in Figure 1). Std. DM effect: fixed-effect regression coefficient divided by the standard deviation of the outcome (from Eq. 2; all models control for classroom fixed effects and include a constant). SE: Robust standard error on Cohen’s D , clustered at school level. P-value: unpenalized p-value. BH rank: Rank the p-values, P_r , within a test set in ascending order $r = 1, \dots, k, \dots, \max(r)$. BH 5% value: $\frac{r}{\max(r)} * 0.05$. BH significant: Test meets the BH standard: Find the largest p-value, $P_k \leq \frac{k}{\max(r)} * 0.05$; reject if $P_r \leq P_k$ (note: in this example, it so happens that all BH significant tests have negative values in column 10, but that’s not always the case.)

Table A6. Multiple-hypothesis testing corrections for Figure 2

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Category	Student	DM	Dependent variable	Std. DM effect	SE	p-value	BH rank (r)	BH 5% value	(7) – (9)	BH Significant
gG	Girl	Girl	Reading test score	0.00	0.04	0.9828	2	0.0250	0.9578	0
gB (ref)	Girl	Boy	Reading test score	0.00						
bG	Boy	Girl	Reading test score	0.03	0.04	0.5051	1	0.0125	0.4926	0
bB (ref)	Boy	Boy	Reading test score	0.00						
gG	Girl	Girl	GPA	0.05	0.02	0.0420	1	0.0125	0.0295	0
gB (ref)	Low	Boy	GPA	0.00						
bG	Boy	Girl	GPA	0.02	0.02	0.4217	2	0.0250	0.3967	0
bB (ref)	Middle	Boy	GPA	0.00						
gG	Girl	Girl	Grammar grade	0.06	0.03	0.0299	1	0.0042	0.0258	0
gB (ref)	Low	Boy	Grammar grade	0.00						
bG	Boy	Girl	Grammar grade	0.04	0.03	0.1860	4	0.0167	0.1693	0
bB (ref)	Middle	Boy	Grammar grade	0.00						
gG	Girl	Girl	Literature grade	0.06	0.03	0.0440	2	0.0083	0.0357	0
gB (ref)	Low	Boy	Literature grade	0.00						
bG	Boy	Girl	Literature grade	0.02	0.03	0.4749	8	0.0333	0.4416	0
bB (ref)	Middle	Boy	Literature grade	0.00						
gG	Girl	Girl	Mathematics grade	0.03	0.03	0.3010	5	0.0208	0.2801	0
gB (ref)	Low	Boy	Mathematics grade	0.00						
bG	Boy	Girl	Mathematics grade	0.00	0.02	0.9301	12	0.0500	0.8801	0
bB (ref)	Middle	Boy	Mathematics grade	0.00						
gG	Girl	Girl	Behavior grade	0.01	0.04	0.7664	10	0.0417	0.7247	0
gB (ref)	Low	Boy	Behavior grade	0.00						
bG	Boy	Girl	Behavior grade	-0.01	0.04	0.8253	11	0.0458	0.7794	0
bB (ref)	Middle	Boy	Behavior grade	0.00						

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Category	Student	DM	Dependent variable	Std. DM effect	SE	p-value	BH rank (r)	BH 5% value	(7) – (9)	BH Significant
gG	Girl	Girl	Diligence grade	0.02	0.02	0.4019	6	0.0250	0.3769	0
gB (ref)	Low	Boy	Diligence grade	0.00						
bG	Boy	Girl	Diligence grade	0.02	0.03	0.4813	9	0.0375	0.4438	0
bB (ref)	Middle	Boy	Diligence grade	0.00						
gG	Girl	Girl	Teacher-written tests	0.02	0.03	0.4339	7	0.0292	0.4047	0
gB (ref)	Low	Boy	Teacher-written tests	0.00						
bG	Boy	Girl	Teacher-written tests	0.04	0.02	0.1329	3	0.0125	0.1204	0
gG	Girl	Girl	Diligence grade	0.02	0.02	0.4019	6	0.0250	0.3769	0

Note: Grey background means „significant after Benjamini-Hochberg (1995) correction for multiple hypothesis testing, with a 5% false-discovery rate.” As in all analyses, we consider the coefficients in each primary-outcome model as its own set (i.e., 4 tests each in panels 1 and 2 of Figure 2), and the collection of deskmate tests across all secondary outcomes another set (12 tests across panels 3-8 in Figure 2). Std. DM effect: fixed-effect regression coefficient divided by the standard deviation of the outcome (from Eq. 3; all models control for classroom fixed effects and include a constant). SE: Robust standard error on Cohen’s D, clustered at school level. P-value: unpenalized p-value. BH rank: Rank the p-values, P_r , within a test set in ascending order $r = 1, \dots, k, \dots, \max(r)$. BH 5% value: $\frac{r}{\max(r)} * 0.05$. BH significant: Test meets the BH standard:

Find the largest p-value, $P_k \leq \frac{k}{\max(r)} * 0.05$; reject if $P_r \leq P_k$. In this table, no estimate remains significant after the BH correction.

Appendix B: National Test Scores as Alternative Exposure and Outcome Measures (Un-preregistered)

After the conclusion of our experiment, we obtained detailed, PISA-like, nationally standardized test scores from the Hungarian National Assessment of Basic Competencies (NABC) for a sample of students. The NABC assesses student achievement in mathematics and reading comprehension in four 45-minute exams at the end of 6th and 8th grade.

NABC scores enable a series of (un-preregistered) robustness checks for our main specifications. Hence, we use recent 6th-grade NABC scores as an alternative measure of baseline achievement-for consenting 7th-grade students, and NABC scores from 6th and 8th grade as an alternative learning outcome measure for consenting 6th-and 8th-grade students in our sample.

Positively, compared to the GPA-based exposure and outcome measures in our primary analysis, machine-graded NABC scores are more detailed/precise and free from potential grader bias. Negatively, approximately 30 percent of students have missing NABC scores,¹ and missingness is selective of having higher baseline achievement.

[TABLE B1 ABOUT HERE]

Using within-sample standardized NABC scores from the end of 6th grade as an alternative measure of baseline achievement-in the analysis of deskmate effects on the learning outcomes of 7th-grade students, we find a substantively large deskmate effect of sitting next to a higher-achievement-deskmate on students' 7th grade mathematics grade, but not on any other primary or secondary outcome (Appendix Table B1). This effect is statistically significant at conventional (unpenalized) 5% level, but it is not statistically significant at the 5% false-discovery rate after correcting for multiple hypothesis testing.

[TABLES B2 AND B3 ABOUT HERE]

¹ NABC scores were missing for a variety of reasons, including students not having participated in the NABC (esp. in 6th grade), missing parental consent to use NABC scores for this study, and missing linkage IDs.

Using NABC scores (in math, reading comprehension, and their average) as outcomes for 6th and 8th graders, we find no statistically significant evidence for an effect of sitting next to a higher-baseline GPA deskmate (Appendix Table B2) or a female deskmate (Appendix Table B3) on students' endline achievement.²

These results thus support the conclusions of the pre-registered analyses (reported in the body of the text) that there is no strong evidence for a positive effect of sitting next to a higher-achievement-student on the average level of students' educational outcomes.

² In 2008, 6th-grade NABC scores were nationally standardized to mean 1500 and SD 200 and are allowed to change across grade levels and over time. This original standardization does not limit our ability to detect deskmate effects on average achievement levels in 2018 in our sample, as our sample is small compared to the population of test takers in Hungary.

Table B.1: Deskmate effects on students' achievement by deskmate's baseline NABC score, 7th-grade students (un-preregistered)

	Primary outcomes		Secondary outcomes					
	(1) Reading score	(2) Endline GPA	(3) Endline Grammar	(4) Endline Literature	(5) Endline Mathematics	(6) Endline Behavior	(7) Endline Diligence	(8) Scores on teacher-written tests
Own baseline NABC score	9.36 [†] (1.25)	0.82 [†] (0.04)	0.77 [†] (0.05)	0.81 [†] (0.07)	0.89 [†] (0.05)	0.41 [†] (0.08)	0.70 [†] (0.05)	15.09 [†] (0.97)
Deskmate's baseline NABC score	0.18 (0.65)	0.08 (0.05)	0.11 (0.07)	0.04 (0.06)	0.11* (0.05)	0.06 (0.04)	0.05 (0.05)	1.23 (1.17)
Female deskmate	1.48 (1.77)	0.34 [†] (0.10)	0.43 [†] (0.10)	0.47 [†] (0.14)	0.10 (0.11)	0.60 [†] (0.10)	0.35 [†] (0.10)	4.56 [†] (1.57)
Classroom fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of the outcome	62.48	3.27	3.34	3.34	3.12	4.06	3.60	59.20
Standardized DM effect	0.01	0.08	0.10	0.03	0.10	0.06	0.05	0.06
Observations	275	300	300	300	297	300	300	252
R-squared	0.33	0.61	0.51	0.48	0.54	0.29	0.46	0.64

Note: NABC score: Average of math and reading-comprehension score on the Hungarian National Assessment of Basic Competencies; available as recent measure of baseline achievement only for 7th graders (test taken at end of 6th grade, 2017); for this analysis only, NABC scores are standardized within the sample of to mean=0, SD=1. GPA is computed from grades in grammar, literature, and mathematics (5 is best, 1 is worst). All models control for classroom fixed effects and include a constant. No additional controls are included. Standardized deskmate effects equal the deskmate coefficient divided by the standard deviation of the outcome. Robust standard errors are clustered at the school level in parentheses. Table reports conventional (unpenalized) hypothesis tests: † p<0.01, * p<0.05. Penalizing significance levels for multiple hypothesis testing by Benjamini and Hochberg's (1995) procedure with a 5% false-discovery rate across deskmate coefficients for the secondary outcomes indicates no statistically significant coefficients.

Table B2: Deskmate effects on students' NABC score by deskmate's baseline GPA, 6th- and 8th-grade students (un-preregistered)

	(1) NABC, average	(2) NABC mathematics	(3) NABC reading comprehensions	(4) NABC mathematics	(5) NABC reading comprehensions	(6) NABC reading comprehensions
Own baseline GPA	114.82† (5.14)	106.40† (5.81)	123.24† (6.09)			
Deskmate's baseline GPA	2.38 (3.98)	8.51 (4.51)	-3.76 (5.05)			
Own baseline mathematics grade				86.24† (5.10)		
Deskmate's baseline mathematics grade				7.84 (4.33)		
Own baseline grammar grade					109.44† (5.13)	
Deskmate's baseline grammar grade					-0.70 (6.08)	
Own baseline literature grade						95.31† (6.69)
Deskmate's baseline literature grade						-1.79 (5.04)
Female deskmate	-17.10 (10.28)	-51.01† (8.83)	16.81 (13.37)	-45.76† (10.82)	11.89 (14.35)	38.63† (13.14)
Classroom fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Mean of the outcome	1486.56	1497.64	1475.48	1498.69	1475.48	1475.48
Standardized DM effect	0.01	0.05	-0.02	0.04	-0.00	-0.01
Observations	655	655	655	646	655	655
R-squared	0.52	0.44	0.48	0.35	0.43	0.38

Note: NABC score: Hungarian National Assessment of Basic Competencies; available as a proximal outcome measure (measured at the end of the spring term) only for 6th- and 8th-graders (test taken in May 2018); GPA is computed from grades in grammar, literature, and mathematics (5 is best, 1 is worst). All models control for classroom fixed effects and include a constant. No additional controls are included. Standardized deskmate effects equal the deskmate coefficient divided by the standard deviation of the outcome. Robust standard errors are clustered at the school level in parentheses. Table reports conventional (unpenalized) hypothesis tests: † p<0.01, * p<0.05. Penalizing significance levels for multiple hypothesis testing by Benjamini and Hochberg's (1995) procedure with a 5% false-discovery rate across deskmate coefficients for the secondary outcomes indicates no statistically significant coefficients.

Table B3: Deskmate Effects on Students' NABC Score by Deskmate's Gender, 6th-and 8th-grade students (un-preregistered)

	(1)	(2)	(3)
	NABC, average	NABC mathematics	NABC reading comprehensions
Female student	-14.58 (10.49)	-48.11 [†] (8.86)	18.95 (13.70)
Female deskmate	5.77 (9.12)	6.19 (8.48)	5.35 (11.16)
Own baseline GPA	114.88 [†] (5.26)	105.80 [†] (5.91)	123.96 [†] (6.02)
Classroom fixed effects	Yes	Yes	Yes
Mean of the outcome for those with a male deskmate	1476.32	1486.29	1466.36
Standardized DM effect	0.03	0.03	0.03
Observations	672	672	672
R-squared	0.52	0.44	0.48

Note: NABC score: Hungarian National Assessment of Basic Competencies; available as a proximal outcome measure (measured at the end of the spring term) only for 6th-and 8th-graders (test taken in May 2018). GPA is computed from grades in grammar, literature, and mathematics (5 is best, 1 is worst). All models control for classroom fixed effects and include a constant. No additional controls are included. Standardized deskmate effects equal the deskmate coefficient divided by the standard deviation of the outcome. Robust standard errors are clustered at the school level in parentheses. Table reports conventional (unpenalized) hypothesis tests: [†] p<0.01, * p<0.05. Penalizing significance levels for multiple hypothesis testing by Benjamini and Hochberg's (1995) procedure with a 5% false-discovery rate across deskmate coefficients for the secondary outcomes indicates no statistically significant coefficients.