

Chapter 5

The Interplay Between Instructional Pace, Skill Externalities, and Student Achievement: An Empirical Assessment



David Kiss

Abstract This article empirically validates a theoretical model from Kiss (B.E. J Econ Anal Policy 17:1–10, 2017) that addresses the transmission channels through which increases in peer achievement levels may affect other students' achievement. In his model, a higher share of better students has two effects: first, weaker students benefit from skill externalities generated by their better classmates. At the same time, however, better students further induce teachers to instruct at a more demanding pace. Based on these two effects, I derive three hypotheses and test them on data on German secondary school students. Empirical findings are consistent with the model's predictions: increases in the share of better classmates (a) are always beneficial for good students, (b) may hurt weak students, and (c) boost weak students' achievement if the extent of interaction between better and weaker students is high. Taken together, these findings suggest that encouraging better and weaker students to interact more could be Pareto-improving.

5.1 Introduction

Peer characteristics are a key factor in parental school choice decisions (Bayer et al., 2007; Black, 1999), and therefore in discussions on both ability tracking (Duflo et al., 2011; Hanushek & Wößmann, 2006) and school competition (Altonji et al., 2015; Rouse, 1998). An important reason for this lies in the positive influence of good classmates on student achievement.¹ Nonetheless, little is known about the causes behind the positive relationship between peer and own achievement levels.

¹This has been shown in numerous empirical studies that carefully account for potential endogeneity issues (see, for a summary, Sacerdote, 2011). These studies suggest that, on average, a one standard deviation increase in peer achievement levels boosts own achievement by 0.1 standard deviations. Throughout this article, the terms ability, (academic) achievement, skills, and

D. Kiss (✉)

Leibniz University Hannover, Hanover, Germany

e-mail: kiss@aoek.uni-hannover.de

© The Author(s) 2023

S. Weinert et al. (eds.), *Education, Competence Development and Career Trajectories*, Methodology of Educational Measurement and Assessment, https://doi.org/10.1007/978-3-031-27007-9_5

113

However, any assessment of the expected consequences of interventions that change the skill composition of classes would require a profound understanding of transmission channels.

So far, the theoretical literature has identified disruptive student behaviour (Lazear, 2001), effort spillovers (Foster & Frijters, 2009; Fruehwirth, 2013), and parental investments in their child's education (Das et al., 2013; Pop-Eleches & Urquiola, 2013) as possible causes. This article empirically validates Kiss (2017), who provides an additional theoretical explanation. In his model, better students induce teachers to instruct their classes at a higher instructional pace. Whereas this is beneficial for good students because of their well-developed learning capabilities, weaker students struggle with a more demanding pace. Therefore, the so-called *pace effect* is positive for better students and negative for weaker students. The model further presumes that better students generate positive skill externalities (*spillover effects*) that enhance the learning potential of their weaker classmates. Consequently, better students improve weaker students' achievement only if (negative) pace effects are offset by (positive) spillover effects.

The most important variable in the theoretical model is the share of higher achieving students in a class, denoted by n . Because classes are populated by only two types of student (better and weaker ones), changes in n are equivalent to changes in the average achievement levels of classes. To see why, consider a class in which a weak student is replaced by a better one. Obviously, both the share of better students n and the average achievement level of the class would increase in this case.

Using longitudinal data on German secondary students, I empirically test the following three hypotheses derived from the theoretical model:

- H1. Better classmates have a positive effect on good students' achievement.
- H2. Better classmates *may* have a negative effect on weak students' achievement.
- H3. Weak students benefit from better classmates if the extent of interaction between student types is high.

All three hypotheses are supported by empirical findings, because the signs of the estimated coefficients are in line with the model's predictions.² For math scores, results further indicate that skill externalities are stronger among students of the same gender. This chapter therefore contributes to the literature in two ways. First, empirical support for H2 challenges the general notion that better peers are always beneficial for students.³ Second, the large heterogeneity in the magnitude of

(test) scores are used interchangeably, and 'peer' is just a synonym for 'classmate'. For better readability, the gender of students is female and that of teachers is male.

²The corresponding estimates tend to have the 'correct' signs; however, they turn out to be non-significant in most cases.

³Most of the theoretical (empirical) studies summarized by Epple and Romano (2011) and Sacerdote (2011) presume (report) positive ability peer effects. One of the few exceptions is research on ordinal rank effects: for example, Cicala et al. (2016) and Elsner and Isphording (2017) find that students experiencing a drop in their relative achievement (compared to their classmates) are more likely to have behavioural problems and lower secondary school graduation rates.

estimated ability peer effects found in the literature may stem (partly) from differences in the extent of interaction between student types.

In addition, I would like to highlight two aspects of this article: first, the purpose of this article is to provide empirical evidence that is primarily descriptive but consistent with the causal relationships outlined in the theoretical model. Second, to empirically detect diverging impacts of the interplay between pace and spillover effects on various student types, the empirical assessments only focus on students with higher or lower skills. According to the theoretical model, the impacts on the “median student” are expected to lie between the more polar cases that are analyzed here.

This chapter proceeds as follows: Sect. 5.2, “Model and hypotheses” briefly summarizes Kiss (2017) and derives the three hypotheses. Section 5.3, “Empirical strategy” presents the data and the empirical strategy. Section 5.5 “Results” reports the results. The final section provides a summary and conclusions.

5.2 Model and Hypotheses

5.2.1 Summary of the Theoretical Model

In Kiss (2017), classes are populated by two student *types*

$$\theta \in \{l, h\}$$

with l types denoting weaker students and h types denoting better students. Each type’s learning capability or *potential* q_θ equals

$$\begin{aligned} q_h &= h \\ q_l &= l + s(n, i) \end{aligned}$$

with $q_h > q_l > 0$. That is, the largest amount of knowledge an h type can accumulate during a period (say, a school year) equals her potential $q_h = h$.

Regarding l types, their learning potential is determined by both their type and a non-negative function $s(n, i)$ that captures the extent of knowledge externalities (‘spillovers’) generated by their better classmates. *Spillovers* $s(n, i)$ are a positive function of two variables: $n \in (0, 1)$ denotes the *share* of h types (i.e. better peers) in a class. $s(n, i)$ is assumed to be increasing in n , because better peers may have a positive effect on l types’ learning efforts or help them through study collaborations. The second variable $i \in (0, 1)$ denotes the extent of *interaction* between h types and l types, with larger values representing higher levels of interaction. Because of that,

$s(n, i)$ is also increasing in i .⁴ If, for example, a student's type θ is correlated with socio-economic status, then i may vary across classes due to regional differences in the extent of social segregation.

Even though spillovers $s(n, i)$ are non-negative, they are further constrained from above, so that $q_h > q_l > 0$ holds for any n and i , meaning that h types have the potential to learn more than l types. The outcome of interest, however, is a student's final *achievement*

$$a_\theta(p) = q_\theta - |p - q_\theta|,$$

which is a function of her potential q_θ and the instructional *pace* p . The pace p is set by the teacher and reflects the amount of material covered during a school year. From this it becomes apparent that

$$a_\theta(p=q_\theta) = q_\theta > a_\theta(p \neq q_\theta). \quad (5.1)$$

Equation 5.1 means that a student can realize her full potential $q_\theta = \max_p a_\theta(p)$ only if the instructional pace is perfectly targeted. Consequently, final achievement a_θ is depressed whenever $p \neq q_\theta$. The intuition behind this is simple: a student cannot realize her full potential whenever she is bored ($p < q_\theta$) or overchallenged ($p > q_\theta$). Therefore, one can think of q_θ as both θ 's learning potential and θ 's *optimal pace*.

5.2.2 The Three Hypotheses

Figure 5.1 plots a_θ for each student type as a function of the share of better students n , while holding the extent of interaction i constant.⁵ One can see that $a_h > a_l$ for any n , meaning that h types learn more than l types. Regarding the marginal effect of increases in n on each type's learning, two things become apparent. First,

H1: The higher the share of better students n , the more h types learn.

This is inferred simply from the fact that an h type's final achievement a_h has a positive slope for any n in Fig. 5.1. The reason for this lies in the endogenous variable p , the instructional pace set by the teacher. Teachers are assumed to choose p based on the following rule: the larger the share of student type θ in a class, the more closely the instructional pace is tailored to that group's potential $q_\theta =$

$\arg \max_p a_\theta(p)$. In the model, teacher utility is therefore maximized at

⁴As will be shown later, changes in n and i affect weak students' achievement in different ways.

⁵All functional form choices (e.g. the teacher's utility function) are stated and discussed in Kiss (2017).

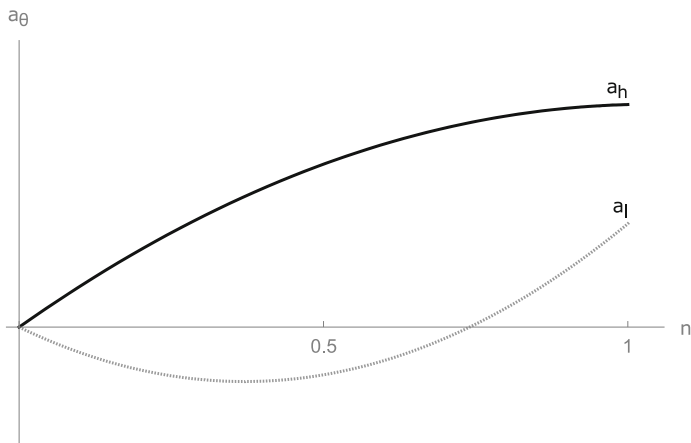


Fig. 5.1 Final achievement a_θ of each student type as a function of n . ($n \in (0, 1)$ denotes the share of better students (h types) in a class. Because there are only two student types $\theta \in \{h, l\}$, higher values of n are equivalent to higher average achievement levels in classes)

$$p^* = n \cdot q_h + (1 - n) \cdot q_l, \tag{5.2}$$

which is a convex combination of each type’s potential q_θ . Obviously, if n was zero or one, then teachers would have chosen a pace that had maximized achievement growth of all students (recall Eq. 5.1 in conjunction with Eq. 5.2). However, as I am interested in mixed classes—which implies $0 < n < 1$ —teachers are assumed to weigh each student type’s optimal pace q_θ by her share in Eq. 5.2.

Because $q_h > q_l$, the teacher’s *optimal pace* p^* is increasing in n —that is, a larger proportion of better students induces teachers to instruct at a more demanding level. In the lingo of the model, the so-called *pace effect* is positive for h types, because a more demanding pace allows h types to better realize their potential.⁶

Second, one can infer from graph a_l in Fig. 5.1 that

H2: The marginal effect of n on an l type’s achievement is negative for small n , and becomes positive once n exceeds some threshold.

The model makes this interesting prediction because increases in the share of better students are accompanied by two opposing effects on l types’ learning: on the one hand, better peers generate additional (positive) knowledge externalities, which is referred to as the *spillover effect*. At the same time, however, better peers further induce teachers to set a higher instructional pace, which turns out to be too demanding for weaker students. Therefore, the net effect of increases in n on a_l depends on the interplay between spillover and pace effects. For small n , positive spillover

⁶ Additionally, note that changes in n and i affect p^* in different ways: both n and i are determinants of $q_l = l + s(n, i)$; however, the weighting of q_h and q_l in the teacher’s utility-maximizing pace p^* depends solely on n .

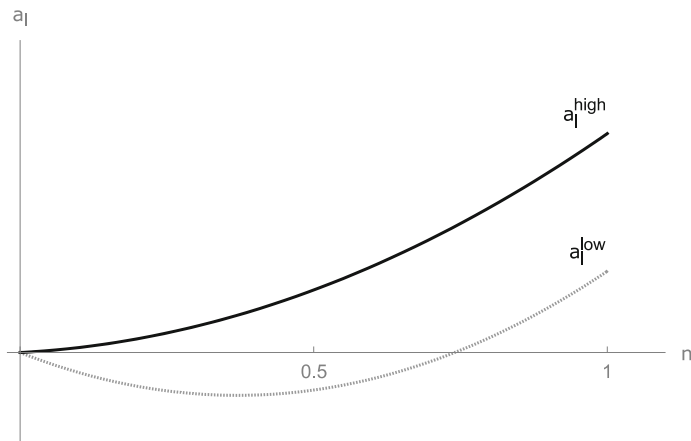


Fig. 5.2 Final achievement of l types (high vs. low levels of interaction). (Increases in the extent of interaction between better and weaker student types are represented by a counterclockwise rotation of the graph of a_l)

effects are dominated by negative pace effects. Once n is sufficiently large, however, further increases in n turn out to generate knowledge externalities that overcompensate negative pace effects.

So far, the extent of interaction between student types was held constant in Fig. 5.1. In Fig. 5.2, achievement of weaker students a_l is plotted twice—once for low and once for high levels of interaction. One can observe that

H3: An increase in the extent of interaction raises l types' achievement.

The reason for this is the following: higher interaction levels generate additional skill externalities that translate into higher achievement levels a_l . Put differently: increases in the extent of interaction are represented in Fig. 5.2 by a counterclockwise rotation of the graph of a_l . Figure 5.2 further suggests that the detrimental effect of better peers for low initial values of n —as stated in H2—may not exist if interaction levels are sufficiently high.

5.3 Empirical Strategy

5.3.1 Data, Institutions, and Descriptive Statistics

This study employs Starting Cohort 3 of the National Educational Panel Study (NEPS). In these data, students are assessed the first time in the autumn of 2010 after enrolling in 5th grade of German secondary school. Thereafter, their educational progress is tracked in yearly follow-up assessments. The key variables used in this study are math skills in Grades 5 and 7: math skills in Grade 5 are used to

classify students into l and h types that then allows me to compute the share of better students n . Math skills at Grade 7 serve as a proxy for the outcome of interest: final achievement a_θ .⁷

Even though NEPS students are also assessed at later grades, this study focuses on only the early stages of a student's secondary school career. The reason lies in the key explanatory variable n that is required to correlate as little as possible with other unobservable determinants of final achievement a_θ . Whereas the analysis presented here is primarily descriptive, biases due to self-selection or omitted variables can be reduced by exploiting the fact that new classes are created at 5th grade—the beginning of German secondary education.

Compared to later grades, potential confounding factors should correlate less with n among 5th graders for the following reasons: (a) students in newly formed 5th-grade classes usually graduated from different elementary schools; (b) parents who are sensitive to peer characteristics require some time to properly assess their child's new class environment; and (c) similarly, teachers also first have to interact with their new classes before they can decide whether, for instance, they are willing to continue instructing them in later grades.⁸

The analysed data are summarized by student type in Table 5.1. Based on the distribution of math scores of all 5th graders in the sample, a student is classified as l type if her 5th grade math score lies below the 25th percentile and as h type if her math score lies above the 75th percentile. As expected, average math percentile scores are much lower for l types than for h types (.12 vs. .87). The next row contains summary statistics for a student's math percentile score at the beginning of 7th grade that serves as the proxy for final achievement a_θ . There is some regression to the mean, because the percentile scores of l types are increasing and those of h types are decreasing over time. The remaining variables are self-explanatory: girls are somewhat over-represented among l types and under-represented among h types. In addition, l types are more likely to have less educated parents and to enrol in lower level secondary schools.

5.3.2 Empirical Tests of H1 and H2

To empirically test the relationship between final achievement and the share of better classmates, the following regression model is estimated:

⁷Unfortunately, math skills are assessed in NEPS only once every 2 years (see Bela et al., 2012, for details and Blossfeld et al., 2011, for a general introduction to the NEPS). However, this should not be a great cause for concern because (a) math skills are highly autocorrelated and (b) the composition of classes is relatively stable over time.

⁸Transitions from primary to secondary education serve as natural experiments in the ability peer effects studies conducted by Gibbons and Telhaj (2016), Kiss (2013), and Sund (2009).

Table 5.1 Descriptive statistics (German secondary students)

Student type	Weak (l type)		Good (h type)	
Variable	M	SD	M	SD
Math percentile (5th grade)	.12	.07	.87	.07
Math percentile (7th grade)	.18	.15	.75	.19
Female (5th grade)	.57		.39	
Age (in years, 5th grade)	10.6	.64	10.3	.53
Parental education				
Lower secondary	.10		.01	
Vocational training	.48		.19	
High school	.08		.08	
Tertiary	.34		.72	
Type of secondary school				
Lower secondary	.28		.01	
Multi-track	.13		.04	
Middle secondary	.30		.09	
Comprehensive	.12		.02	
Upper secondary	.17		.84	
N(students)	580		768	

Standard deviations not reported for dummy variables. Unit of analysis: secondary students who were enrolled in 5th grade when first assessed. A student is classified as *l* type below the 25th percentile in the math achievement distribution at the beginning of the 5th grade and as *h* type if her math score lies above the 75th percentile. ‘Parental education’ denotes the highest educational degree of a student’s parents. It is classified according to the International Standard Classification of Education (ISCED) guidelines, see OECD (1999) for details

$$mpc_{i7} = \beta \cdot n_{c5} + \gamma \cdot n_{c5}^2 + \delta' \cdot \mathbf{x}_{i5} + \varepsilon_{ic5} \quad (5.3)$$

in which $mpc_{i7} \in [0, 1]$ denotes the math achievement percentile of student *i* at the beginning of 7th grade, and serves as a proxy for final achievement a_θ . The variable of interest is n_{c5} , the share of higher achieving classmates in Grade 5. The *c* subscript indicates that n_{c5} is constant across students attending the same class. As shown in Fig. 5.1, the relationship between math skills and *n* is nonlinear—to account for this, Eq. 5.3 further includes a square of n_{c5} . \mathbf{x}_{i5} denotes a vector of control variables (age, gender, indicators for parental education, and school type). Errors ε_{ic5} are clustered on the class level.

H1 is tested by first restricting the data to *h* types and then estimating Eq. 5.3. According to the theoretical model, returns to increases in *n* are positive, but diminishing for *h* types (see Fig. 5.1). Because

$$\frac{\partial mpc_{i7}}{\partial n_{c5}} = \beta + 2\gamma \cdot n_{c5}, \quad (5.4)$$

one would therefore expect $\hat{\beta} > 0$ and $\hat{\gamma} < 0$.

H2 is tested in the same fashion: the data are now constrained to l types before Eq. 5.3 is estimated. According to Fig. 5.1, $\frac{\partial mpc_{l7}}{\partial n_{c5}}$ is negative for small n and eventually becomes positive as n further increases—this is the case if $\hat{\beta} < 0$ and $\hat{\gamma} > 0$ in Eq. 5.4.

5.3.3 The Empirical Test of H3

It was stated in Sect. 5.2, “Model and hypotheses” that the extent of interaction may vary across classes due to regional differences in social segregation levels. The empirical test of H3, however, is based on the following assumption:

A1. Same-gender peers interact more than different-gender peers.

To exemplify A1, consider the following 5th-grade class of 22 students: 6 high-achieving girls, 5 high-achieving boys, 6 low-achieving girls, and 5 low-achieving boys. What would be the marginal effect of an additional high-achieving girl on the achievement levels of low-achieving girls and boys? Under A1, the positive effect of that additional high-achieving girl would be stronger on low-achieving girls than on low-achieving boys, because same-sex friendships and social ties are both more likely to develop and be sustained over time.

Under A1, the following procedure is implemented to test H3:

1. Based on a percentile-threshold π , each student (regardless of gender) is classified as *either h or l* type.⁹
2. Each class c is split into two subclasses $c1$ and $c2$. This is done in two different ways, called *scenarios*:
 - Scenario 1: The first subclass $c1$ consists of all high- and low-achieving girls who are enrolled in c . The second subclass $c2$ comprises all high- and low-achieving boys enrolled in c .
 - Scenario 2: $c1$ comprises all high-achieving girls and low-achieving boys enrolled in c . Similarly, all high-achieving boys and low-achieving girls in c are grouped into subclass $c2$.
3. For both scenarios (same- vs mixed-sex subclasses), the share of better students in each subclass is computed. To be precise,

⁹Table 4.1, for example, π is set to 25 to identify l types and to 75 to identify h types.

- n either equals $n_{c1,5}^{gg}$, the share of high-achieving girls in $c1$, the ‘girls-only’ subclass (in 5th grade). Alternatively, $n = n_{c2,5}^{bb}$ equals the share of high-achieving boys in the ‘boys-only’ subclass $c2$.
- Under the second scenario, n equals either $n_{c1,5}^{gb}$ or $n_{c2,5}^{bg}$.¹⁰

4. Estimate Eq. 5.3 and treat subclasses $c1$ and $c2$ as if they were separate classes

This estimation procedure has the appealing property of using the same data to construct subclasses in two different ways. A comparison of estimates therefore allows inferences to be drawn on the effect of changes in the extent of interaction on the achievement growth of l types: because the same data are used in both scenarios, differences in results can emerge only from differences in the way subclasses were constructed. Under A1, the extent of interaction is higher in same-sex subclasses (see H3 and Fig. 5.2). In terms of marginal effects (see Eq. 5.4 in conjunction with H2), these considerations translate into $\hat{\beta}^{low} < \hat{\beta}^{high}$ and $\hat{\gamma}^{low} > \hat{\gamma}^{high}$, with ‘low’ indicating whether the analysis was based on the mixed-sex sample in which I assume interaction levels to be low between student types, and ‘high’ indicating whether the analysis was based on the same-sex sample in which I assume interaction levels to be high between student types.

5.4 Results

5.4.1 Main Findings

H1 is tested with the following three-step procedure:

1. Based on a percentile threshold $\pi \in \{75, \dots, 85\}$, the h type dummy is coded as follows:

$$h \text{ type}_{i5} = \begin{cases} 1 & \text{if } mpc_{i5} \geq \pi, \\ 0 & \text{if } mpc_{i5} < \pi \end{cases}, \text{ that is, } h \text{ type}_{i5} = 1 \text{ if } i\text{'s math skill (measured at the beginning of the 5th grade) is equal or greater than the } \pi^{\text{th}} \text{ percentile. Otherwise, } h \text{ type}_{i5} = 0 \text{ for any student } i \text{ scoring below } \pi.$$

2. Based on $h \text{ type}_{i5}$, the share of high achievers n_{c5} is computed for each class c .
3. Before Eq. 5.3 is estimated, the data are restricted to h types only.

Point estimates of β and γ are plotted against π in Fig. 5.3: for any $75 \leq \pi \leq 85$, estimates of β turn out to be positive—that is, the larger the share of high-achieving students in a class, the more h types learn. However, the graph of a_h is only concave for negative values of γ (recall Eq. 5.4). Because $\hat{\gamma}$ turns out to be positive in some cases, Fig. 5.3. confirms H1 only partially—it should be noted, however, that the

¹⁰In our hypothetical class of 22 students (before the high-achieving girl was added), these shares are $n_{c1,5}^{gg} = \frac{6}{12}$, $n_{c2,5}^{bb} = \frac{5}{10}$, $n_{c1,5}^{gb} = \frac{6}{11}$, and $n_{c2,5}^{bg} = \frac{5}{11}$.

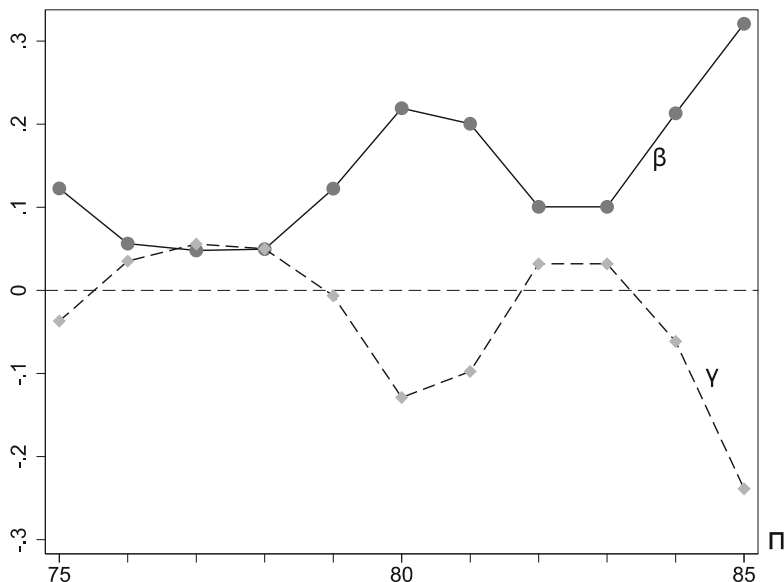


Fig. 5.3 Empirical test of H1 (estimates of β and γ for h types). (This figure reports estimates of β (solid line) and γ (dashed line) for h types for various thresholds $\pi \in \{75, \dots, 85\}$. H1 is derived from the theoretical model that predicts $\hat{\beta} > 0$ and $\hat{\gamma} < 0$ for h types. Sample sizes range from 449 ($\pi = 85$) to 768 ($\pi = 75$). Both the estimation procedure and the relationship between β and γ are discussed at the beginning of Sect. 5.5, “Results”)

sign of $\hat{\gamma}$ is of secondary importance with respect to the theoretical prediction that final achievement of h types is increasing in n . The average values of $\hat{\beta}$ and $\hat{\gamma}$ are .14 and $-.03$, respectively, and non-significant for any $\pi \in \{75, \dots, 85\}$.¹¹

H2 is tested in the same manner. The l type dummy

$$l \text{ type}_{i5} = 1(\text{mpc}_{i5} \leq \pi) \text{ for } \pi \in \{15, \dots, 25\}$$

identifies students who score at the π th percentile or below, and π now takes on values between 15 and 25. The share of l types is computed for each class and interpreted as $1 - n_{c5}$. Similar to the empirical test of H1, the data are restricted to l types before Eq. 5.3 is estimated.

¹¹One might be puzzled by the fact that the graphs of β and γ are symmetric along some horizontal axis. Recalling Fig. 4.1 in conjunction with Eq. 5.3, note that any specific trajectory in Fig. 4.1 is represented by a unique combination of β and γ . Therefore, similar values of $\hat{\beta}$ and $\hat{\gamma}$ under two different π imply that the graphs of a_h are (almost) overlapping. The symmetry arises for the following reason: because $\text{mpc}_7 \in [0, 1]$ is bounded from above, the curvature of ‘initially steep’ graphs of a_h (that is, for large $\hat{\beta}$) must be more pronounced, which implies a large absolute value of $\hat{\gamma}$. Put differently, the degree of curvature of the positively sloped, concave, and bounded graph a_h imposes a mechanical relationship between β and γ .

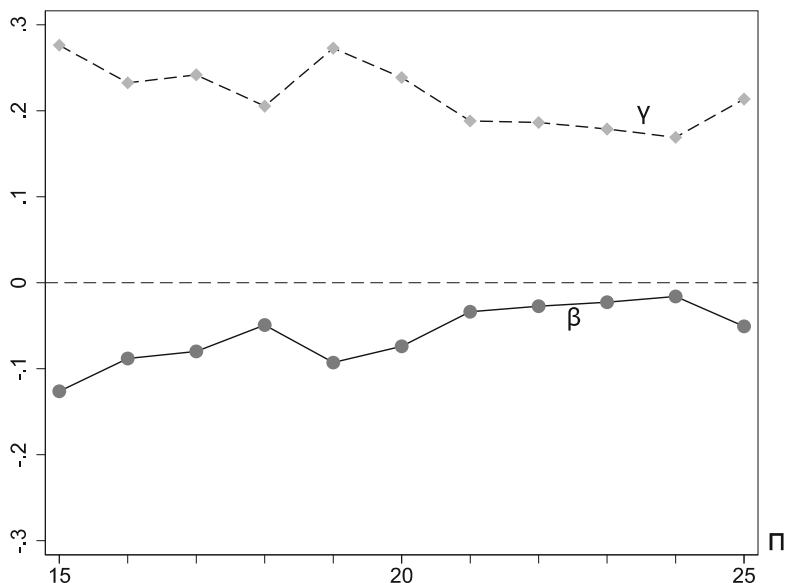


Fig. 5.4 Empirical test of H2 (estimates of β and γ for l types). (This figure reports estimates of β (solid line) and γ (dashed line) for l types under various thresholds $\pi \in \{15, \dots, 25\}$. Here, the share of l types in a class is interpreted as $1 - n$. H2 is derived from the theoretical model, that predicts $\hat{\beta} < 0$ and $\hat{\gamma} > 0$ for l types. Sample sizes range from 334 ($\pi = 15$) to 580 ($\pi = 25$))

Point estimates of β and γ are plotted against π in Fig. 5.4. Consistent with H2, point estimates of β are now negative and point estimates of γ are now positive for any π , and the average value of $\hat{\beta}$ is .06 and of $\hat{\gamma}$ is .21. All estimates of β turn out to be non-significant and all estimates of γ to be significant. It should further be noted that the point estimates of β become positive if $\pi > 30$, i.e., if the threshold to identify l -types takes on “too large” values.¹²

To test H3, each class c is split into subclasses $c1$ and $c2$ in two different ways (see Sect. 5.3.3, “The empirical test of H3” for details). In both cases, each subclass encompasses h and l types. Under the first scenario, $c1$ contains only girls and $c2$ only boys. In the second scenario, high-achieving girls and low-achieving boys constitute $c1$, and subclass $c2$ comprises high-achieving boys and low-achieving girls. The marginal effect of increases in n should therefore depend on the way subclasses are constructed (same-sex vs mixed-sex). Based on the theoretical model, one would expect larger marginal effects in classes with higher levels of interaction (see Eq. 5.4). In terms of regression coefficients on the marginal effect of increases in the share of better classmates on l types, this notion translates into $\hat{\beta}^{\text{high}} > \hat{\beta}^{\text{low}}$ and $\hat{\gamma}^{\text{high}} < \hat{\gamma}^{\text{low}}$ —with ‘high’ indicating whether the analysis was conducted on the

¹²Put differently: to empirically assess the model’s predictions, there must be a sufficiently large skill gap between l and h types.

same-sex sample and ‘low’ indicating whether it was conducted on the mixed-sex sample.

Estimates of Eq. 5.3 are plotted in Fig. 5.5 for same-sex (top figure) versus mixed-sex subclasses (bottom figure). One can see that $\hat{\beta}^{\text{high}} > \hat{\beta}^{\text{low}}$ for any threshold $\pi \in \{15, \dots, 25\}$ at which students are classified into l and h types. The average values of $\hat{\beta}^{\text{high}}$ are -0.01 , and the mean values of $\hat{\gamma}^{\text{high}}$ are 0.10 . The average values of $\hat{\beta}^{\text{low}}$ are -0.08 , and the mean values of $\hat{\gamma}^{\text{low}}$ are 0.21 .¹³ In essence, this set of findings resembles the two graphs of a_l in Fig. 5.2, therefore empirically supporting H3: this can be inferred from a_l^{high} being steeper than a_l^{low} , which is reflected by $\hat{\beta}^{\text{high}}$ being larger than $\hat{\beta}^{\text{low}}$. In addition, both a_l^{high} and a_l^{low} are convex, which is also confirmed by the data because $\hat{\gamma} > 0$ in both scenarios.

5.4.2 Robustness Checks Based on Reading Scores

To check the robustness of the results, the analyses from Sect. 5.5.1, ‘‘Main findings’’ are repeated with reading (rather than math) scores—that is, students are classified into l and h types based on their reading scores at the beginning of 5th grade, and final achievement is now proxied by reading scores measured in Grade 7.

The empirical findings reported in Fig. 5.6 are mostly in line with the theoretical model: $\hat{\beta}$ is positive among h types (top figure) and negative for l types (bottom figure). As before, H1 is confirmed only partially, because $\hat{\gamma}$ is negative in only 50% of cases (top figure). However, as implied by H2, $\hat{\gamma}$ is positive among l types for any $\pi \in \{15, \dots, 25\}$. Again, most of the estimated β and γ are non-significant.¹⁴

H3, however, is not empirically supported for reading scores. According to H3, $\hat{\beta}$ should be larger in high-interaction (i.e. same-sex) subclasses. This is not confirmed by the data because $\text{mean}(\hat{\beta}^{\text{high}}) \approx -0.08 < \text{mean}(\hat{\beta}^{\text{low}}) \approx -0.06$. One may wonder whether this finding invalidates H3 or could be explained by some systematic differences in the acquisition of math and reading skills. It might be the case, for instance, that math skills are more easily transferable between student types who are collaborating in study groups. In addition, weaker students (or their parents) may have a greater awareness of both their math deficits and the need to overcome them.

¹³Regarding significance levels, none of the $\hat{\beta}$ and 59% of the $\hat{\gamma}$ are significant.

¹⁴In the top figure, only 9% of the $\hat{\beta}$ and none of the $\hat{\gamma}$ are significant at the 10% level; in the bottom figure, none of the $\hat{\beta}$ and 27% of the $\hat{\gamma}$ are significant.

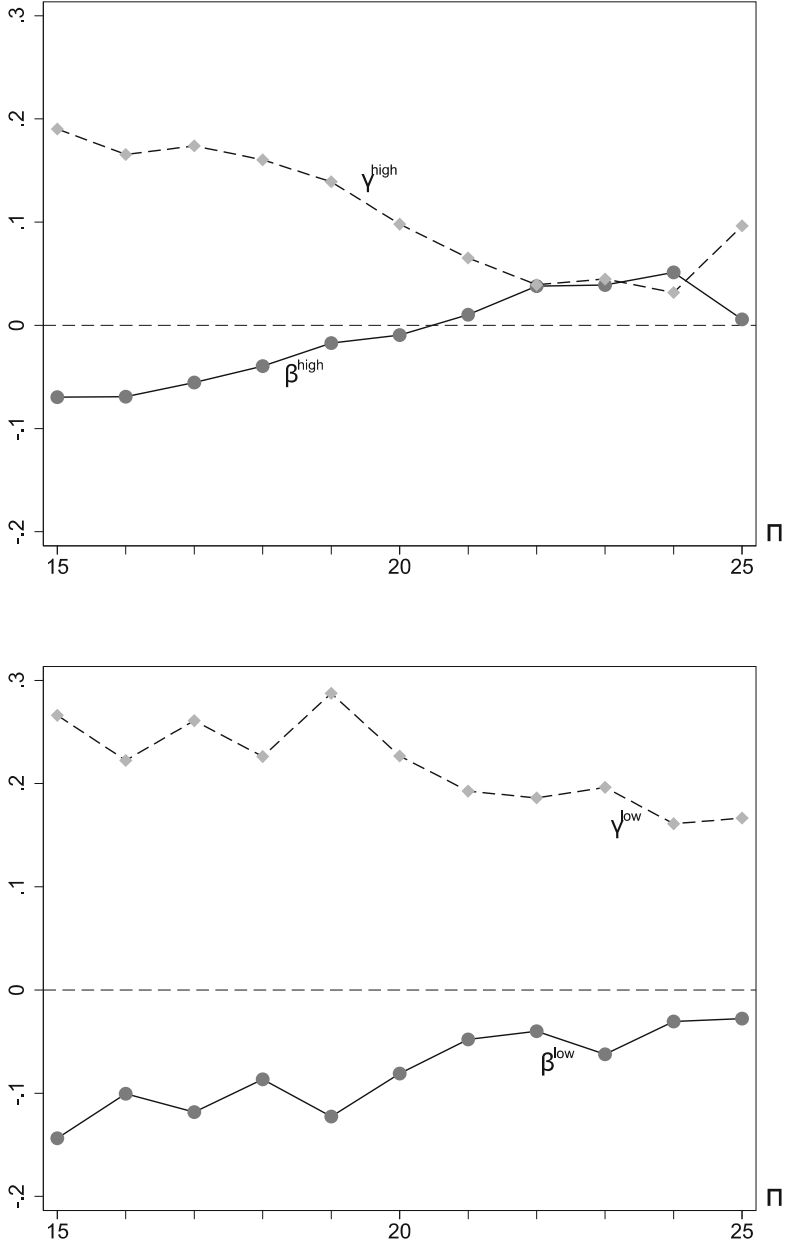


Fig. 5.5 Empirical test of H3 (estimated β and γ in high- and low-interaction classes). (H3 translates into $\hat{\beta}^{\text{high}} > \hat{\beta}^{\text{low}}$ and $\hat{\gamma}^{\text{high}} < \hat{\gamma}^{\text{low}}$ —that is, the marginal effect of better peers is increasing in the extent of interaction between student types. Interaction levels are assumed to be higher in same-sex subclasses (top figure) than in mixed-sex subclasses (bottom figure), see Sect. 5.3.3, “The empirical test of H3” for details)

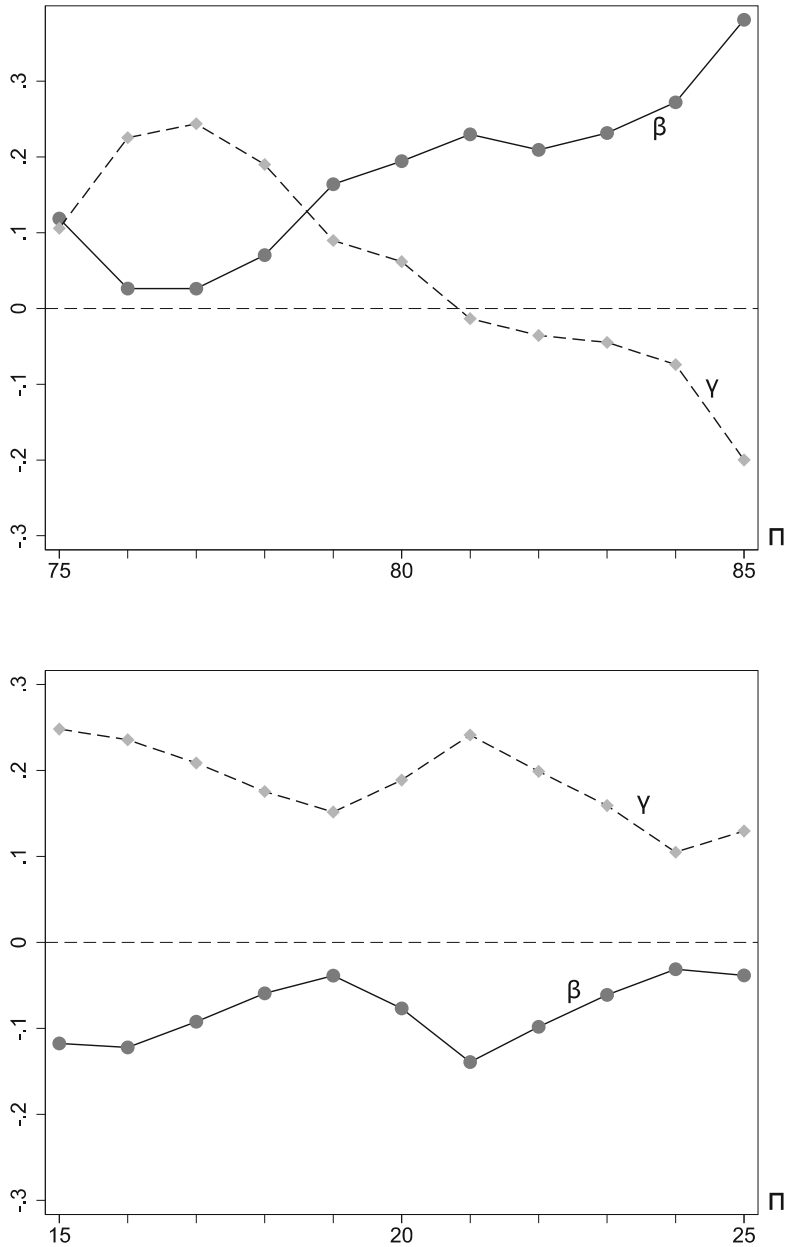


Fig. 5.6 Robustness checks for H1 and H2 based on reading scores. (Unlike Figs. 5.3 and 5.4, students are now classified as l or h types based on their reading scores at the beginning of 5th grade, and final achievement a_θ is proxied by reading scores at the beginning of 7th grade. The top figure reports estimates of β (solid line) and γ (dashed line) for h types under various thresholds $\pi \in \{75, \dots, 85\}$, therefore providing an empirical test for H1. The bottom figure tests H2 by reporting estimates of β and γ for l types for $\pi \in \{15, \dots, 25\}$. Sample sizes range between 445 and 741 in the top figure and between 328 and 589 in the bottom figure)

5.5 Summary and Conclusions

Though there is compelling empirical evidence that students tend to learn more in classrooms with higher shares of good peers, little is still known about the causes behind this relationship. This article complements our understanding about transmission channels by empirically validating the theoretical model on the interplay between instructional pace, skill externalities, and student achievement formulated by Kiss (2017). Three hypotheses on the impact of better peers on student achievement are tested with data on German secondary students. To minimize biases, all analyses are based on a sample of newly formed 5th-grade classes at the beginning of German secondary education.

The empirical findings for math achievement support each hypothesis. Better peers (a) boost good students' achievement, (b) can have a detrimental effect on weak students (presumably because they induce teachers to set a too demanding instructional pace), and (c) raise weaker students' achievement if the extent of interaction between student types is high. Results (a) and (b) are confirmed by robustness checks based on reading scores. However, the third hypothesis is empirically supported only for math scores—one plausible explanation might be the greater awareness of weaker students for both their math deficits and the need to overcome them.

A more profound understanding of transmission channels allows a better assessment of the expected consequences of interventions that change the skill composition of classes. In addition, it may help to identify potentially Pareto-improving interventions—in our case, all students may benefit from learning environments that encourage them to interact more.

References

- Altonji, J., Huang, C.-I., & Taber, C. (2015). Estimating the cream skimming effect of school choice. *Journal of Political Economy*, 123(2), 266–324.
- Bayer, P., Ferreira, F., & McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy*, 115(4), 588–638.
- Bela, D., Pink, S., & Skopek, J. (2012). *Data manual. Starting cohort 3 – From lower to upper secondary school. neps sc3 1.0.0*. Research data article. NEPS, University of Bamberg.
- Black, S. (1999). Do better schools matter? Parental valuation of elementary education. *Quarterly Journal of Economics*, 114(2), 577–599.
- Blossfeld, H.-P., Rossbach, H.-G., & von Maurice, J. (2011). Education as a lifelong process. The German National Educational Panel Study (NEPS). *Zeitschrift für Erziehungswissenschaft, Sonderheft*, 14, 19–34.
- Cicala, S., Fryer, R. G., Jr., & Spenkuch, J. (2016). *Self-selection and comparative advantage in social interactions*. Retrieved February 1, 2018, from https://articles.ssrn.com/sol3/articles.cfm?abstract_id=1780652roprofilöe.com
- Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., & Sundararaman, V. (2013). School inputs, household substitution, and test scores. *American Economic Journal: Applied Economics*, 5(2), 29–57.

- Duflo, E., Dupas, P., & Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review*, *101*(5), 1739–1774.
- Elsner, B., & Ispording, I. (2017). A big fish in a small pond: Ability rank and human capital investment. *Journal of Labor Economics*, *35*(3), 787–828.
- Epple, D., & Romano, R. (2011). Peer effects in education: A survey of the theory and evidence. In J. Benhabib, A. Bisin, & M. O. Jackson (Eds.), *The handbook of social economics* (Vol. 1, pp. 1053–1163). Elsevier.
- Foster, G., & Frijters, P. (2009). *Are there complementarities in educational peer effects?* Retrieved May 3, 2021, from http://research.economics.unsw.edu.au/gfoster/peer_welfare_max_20.pdf
- Fruehwirth, J. (2013). Identifying peer achievement spillovers: Implications for desegregation and the achievement gap. *Quantitative Economics*, *4*(1), 85–124.
- Gibbons, S., & Telhaj, S. (2016). Peer effects: Evidence from secondary school transition in England. *Oxford Bulletin of Economics and Statistics*, *78*(4), 548–575.
- Hanushek, E., & Wößmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *Economic Journal*, *116*(510), 63–76.
- Kiss, D. (2013). The impact of peer achievement and peer heterogeneity on own achievement growth: Evidence from school transitions. *Economics of Education Review*, *37*, 58–65.
- Kiss, D. (2017). A model about the impact of ability grouping on student achievement. *B.E. Journal of Economic Analysis & Policy*, *17*(3), 1–10.
- Lazear, E. (2001). Educational production. *Quarterly Journal of Economics*, *116*(3), 777–803.
- OECD. (1999). *Classifying educational programmes: Manual for ISCED-97 implementation in OECD countries*. OECD Publishing.
- Pop-Eleches, C., & Urquiola, M. (2013). Going to a better school: Effects and behavioral responses. *American Economic Review*, *103*(4), 1289–1324.
- Rouse, C. E. (1998). Private school vouchers and student achievement: An evaluation of the Milwaukee parental choice program. *Quarterly Journal of Economics*, *113*(2), 553–602.
- Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? In E. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (Vol. 3, pp. 249–277). Elsevier.
- Sund, K. (2009). Estimating peer effects in Swedish high school using school, teacher, and student fixed effects. *Economics of Education Review*, *28*(3), 329–336.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

