David Kiss[1]

A Model about the Impact of Ability Grouping on Student Achievement

[1] Leibniz University Hannover, Institute of Labor Economics, Koenigsworther Platz 1, 30167 Hannover, Germany, E-mail: kiss@aoek.uni-hannover.de. http://orcid.org/0000-0002-2475-8532.

Abstract:
This paper presents a small theoretical model to compare school systems that segregate students by ability ("tracking") with comprehensive ones, which allow for mixing of differently skilled students into same classes. The outcomes of interest are the achievement levels of weaker and better students, and the average achievement of all students. In the model, the instructional pace is tailored to the skill distribution of a class, and higher-achieving peers are an additional source of learning. The results show that differences in both the share of high-achievers and degree of interaction between student types can explain the mixed (quasi-)experimental evidence on the effect of de-tracking on student achievement. As changes in peer quality affect good and weak students' achievement in very different ways, the term "peer effect" should be used with caution.

Keywords: decomposition of ability peer effects, tracking vs. mixing
JEL classification: I24, J24, H52
DOI: 10.1515/bejeap-2017-0036

1   Introduction

This paper presents a small theoretical model to investigate the impact of de-tracking on achievement levels of low-achievers, high-achievers, and the entire student body.[1] Proponents of tracking argue that grouping students by ability allows teachers to match their instruction more closely to their students' learning capabilities which benefits both good and weak students: good students are not slowed down by under-performing classmates, and teachers must not worry about losing weaker students. Opponents point out that tracking prevents weaker students from interacting with their higher-achieving peers, who might provide help and serve as role models. As a consequence, tracking may preserve economic inequalities.

(Quasi-)experimental empirical evidence on the impact of de-tracking is inconclusive, see Table 1. For example, results from a tracking experiment in Kenya (Duflo, Dupas, and Kremer 2011) suggest that both lower- and higher-achieving pupils are better off in tracked systems. Somewhat similar to that, Card and Giuliano (2016) report that ability-tracking is beneficial for students enrolled in the upper track without having negative consequences for those remaining in the lower-level track. In both studies, therefore, average achievement is maximized if students are segregated by ability. On the other hand, Pekkarinen, Uusitalo, and Kerr (2013) find that de-tracking raises average achievement in their sample of Finnish secondary students.[2] One may wonder why some studies find positive while others report negative de-tracking effects. Obviously, de-tracking increases the within-class heterogeneity in achievement levels. Though it is much less clear *how* de-tracking affects the educational progress of various student types. The model presented here assumes that increases in the share of good students induce teachers to instruct at a more demanding level (pace effect). It is further assumed that better students create skill-externalities which have a positive effect on weaker students' learning (spillover effect).

Table 1: De-tracking effects in empirical and theoretical studies.

Study	Country	Impact of de-tracking on...					
		Low-achievers		High-achievers		Mean achievement	
Empirical (tracking experiments)							
Lovell (1960)	USA	–		– –	***	–	**
Duflo, Dupas, and Kremer (2011)	Kenya	–	**	–	**	–	**
Card and Giuliano (2016)	USA	none		–	***	–	n.a.
Marascuilo and Maryellen (1972)	USA	++	**	–		+	

Pekkarinen, Uusitalo, and Kerr (2013)	Finland	++	**	+		+	**
Theoretical							
Hidalgo-Hidalgo (2011)		+		−		−	
Lazear (2001)		+		−		−	
Arnott and Rowse (1987)		+		−		±	
This paper		±		−		±	
Meier (2004)		±		±		±	

This table reports signs and significance levels of estimated de-tracking effects on low-achievers, high-achievers, and mean achievement levels based on (quasi-)experimental empirical studies. "De-tracking" refers to a scenario where ability-tracked school systems are replaced by comprehensive ones. A + (−) sign indicates a positive (negative) de-tracking effect at the *10%, **5%, or ***1% significance levels. A doubled sign (e.g., ++) means that the effect is strong compared to other outcomes in *the same* study. Depending on the values of the exogenous parameters, the theoretical models in the bottom panel can produce a range of de-tracking effects. ± means that a model can generate either negative, none, or positive de-tracking effects.

This paper provides four findings. First, as the current instructional pace is too low (high) for better (weaker) students, the pace effect turns out to be positive (negative) for better (weaker) students. Weaker students, however, additionally benefit from spillover effects – the net impact of better classmates therefore depends on whether the negative pace effect is offset by the positive spillover effect. This result suggests that the term "peer effect" should be used with caution.

Second, good students are better off in ability-tracked than in comprehensive school systems, which is in line with most of the empirical evidence. Third, lower-achieving students also prefer tracking if both the share of better students and the extent of interaction between student types are small. Fourth, mixing can maximize average achievement if both the share of good students and extent of interaction are high.

This paper contributes to the literature in two ways. First, the results suggest that differences in both the share of high-achievers and degree of interaction between higher- and lower-achieving students may explain why some empirical studies in Table 1 find positive de-tracking effects while others report negative ones. According to the model, one should expect negative de-tracking effects if both the share of high-achievers and the extent of interaction between student types are low. In that case, a weak student's learning potential is only marginally enhanced by her better peers.[3] At the same time, however, the presence of a few good students induces teachers to instruct at a more demanding level which results in an instructional pace that is too high (low) for weak (good) students, therefore making both types worse off in a mixed system. Mixing can maximize average achievement levels if many well-interacting high-achievers sufficiently increase the learning potential of their weaker peers.

Second, to the best of my knowledge, this paper is the first that decomposes "peer effects" into a pace and spillover effect. The few theoretical papers on de-tracking, which are also summarized in Table 1, take peer effects into account by augmenting a student's educational production function with the mean achievement level of her peers – a change in peer achievement levels is then referred to as "the" peer effect without further distinction. As already mentioned, however, increases in the share of better students may affect higher- and lower-achieving students in quite different ways.[4] The remainder of the paper is as follows. The model is derived in 2. Comparative statics analyzes are presented in 3. 4 contains sensitivity analyzes with regard to the functional form choices made in 2. 5 concludes.

2 The Model

2.1 Determinants of Final Achievement

There are two student types

$$\theta \in \{l, h\} = \{\text{low, high}\}$$

with $h > l > 0$. A student's learning capability or potential

$$p_\theta \equiv \theta + s_\theta \qquad [1]$$

is determined by two factors: her type θ and the extent of spillovers

$$s_\theta \equiv n \, i \, (h - \theta)$$

which is a function of the exogenous variables n, i, h and θ. $n \in [0, 1]$ is the share of h-types in a class. $i \in [0, 1]$ denotes the extent of interaction between h-types and l-types, and is discussed below in more detail.

The definition of s_θ comprises two implicit assumptions. First, only l-types are assumed to benefit from spillovers because $s_h = 0$ but $s_l = n\,i\,(h-l) \geq 0$. For example, l-types may benefit from h-types trough collaborations/study partnerships (Carrell, Fullerton, and West 2009; Arcidiacono et al. 2012; Jain and Kapoor 2015). Exposure to h-types may also have a positive effect on l-types' effort levels (Eisenkopf 2010, Foster and Frijters 2010, and Bursztyn and Jensen 2015). Generally speaking, h-types may serve as role models.[5]

Second, $s_l > 0$ if (i) the class contains at least one h-type (i.e., $n > 0$) who (ii) is actually willing to interact with her weaker peers ($i > 0$). Depending on how social groups are formed, the extent of interaction i may vary across classes. If, for example, social group formation is mainly determined by gender, then lower-achieving boys (girls) may easily establish friendships with their higher-achieving male (female) peers, which would be reflected by a high value of i (Whitmore 2005 and Eisenkopf et al. 2015). In the context of black and white students, however, one would expect i to be small because reference groups are often race-based and black students tend to have lower achievement levels (Hoxby 2000 and Fruehwirth 2013).[6]

Evaluation of eq. (1) for each type yields

$$
\begin{aligned}
p_h &= h \\
p_l &= l + n\,i\,(h-l).
\end{aligned}
\tag{2}
$$

$p_h \geq p_l$ even though h-types do not receive any spillovers. As long as $n < 1$ or $i < 1$, an h-type's (learning) potential exceeds the potential of an l-type. Both $\frac{\partial p_l}{\partial n}$ and $\frac{\partial p_l}{\partial i}$ are positive, i.e., an l-type's learning potential is increasing in n and i as both variables affect s_l positively.

In this study, the outcome of interest is a student's (final) achievement

$$
a_\theta(p) \equiv p_\theta - |p - p_\theta|
\tag{3}
$$

which is a function of her potential p_θ and the (instructional) pace p. The pace reflects the amount of material covered during a school year, and is set by the teacher.[7] One can see that

$$
a_\theta(p = p_\theta) = p_\theta > a_\theta(p \neq p_\theta)
\tag{4}
$$

that is, the highest achievement level type θ can reach is her potential $p_\theta = \max\{a_\theta\}$. However, a_θ is depressed whenever the teacher's pace deviates from θ's potential. Intuitively, $p \neq p_\theta$ means that the student cannot develop her full potential because of either being over-challenged ($p > p_\theta$) or bored ($p < p_\theta$). The larger the mismatch $|p - p_\theta|$, the more $a_\theta(p)$ is depressed. As $a_\theta(p)$ is maximized at $p = p_\theta$, one can interpret p_θ as both θ's potential and optimal pace.

2.2 Teacher's Choice of the Instructional Pace p

Instruction is assumed to be teacher-centered, i.e., teachers never split their time to exclusively instruct subpopulations of their classes. Therefore p is the same for all students. Because (i) h-types' potential exceeds that of l-types, see (2), and (ii) each type's achievement is maximized at $p = p_\theta$ (see eq. (4)), teachers cannot choose a pace that maximizes achievement of both types at the same time. It is therefore assumed that teachers are trying their best to "match" their instructional pace to the skill distribution in their classes. These considerations are modeled as follows. Let

$$
m_\theta(p) \in [0, 1]
$$

denote the (quality of the) match between θ's optimal pace p_θ and the actual pace p chosen by the teacher. $m_\theta(p)$ lies in the unit interval with $m_\theta(p = p_\theta) = 1$ denoting a perfect match which is only realized if $p = p_\theta$. Consequently, because teachers value achievement gains of any student type, their pace must be constrained to values between p_l and p_h. These notions impose the following structure on $m_\theta(p)$:

$$
m_\theta(p) = \begin{cases} 1 & p = p_\theta \\ 0 & p = p_{-\theta} \end{cases}.
$$

That is, $m_h(p_h) = 1$ and $m_h(p_l) = 0$. Thus, as p increases, $m_h(p)$ will also increase because p is approaching p_h from below. The same logic applies to the quality of the match for l-types: because $m_l(p_h) = 0$ and $m_l(p_l) = 1$, the quality of the match for l-types $m_l(p)$ will improve as p decreases.

As already mentioned, teachers cannot maximize both types' achievement in mixed classes because $p_l \neq p_h$ whenever $0 < n < 1$. Therefore, when choosing p, it is assumed that the best teachers can do is to weight each type's match by its share. This motivates the following Cobb–Douglas representation of teacher preferences:

$$
u(p) \equiv m_h(p)^n \cdot m_l(p)^{1-n}.
\tag{5}
$$

The teacher's only choice variable – the instructional pace p – maps into both, the quality of the match for h-types and l-types. As already shown, increases in p benefit h-types but hurt l-types, and vice versa. Teachers account for this trade-off by weighting each type's match by its share: $m_h(p)$ is weighted by the share of high-achievers n, and $m_l(p)$ by the share of low-achievers $1-n$.

To solve eq. (5) for the teacher's optimal pace p^* at which $u(p)$ is maximized, the functional form of $m_\theta(p)$ has to be specified. The simplest choice is

$$m_\theta(p) \equiv \frac{|p - p_{-\theta}|}{p_h - p_l},$$

which yields $m_h(p) = \frac{p - p_l}{p_h - p_l}$ for h-types. This function possesses all required properties: (i) $m_h(p)$ lies in the unit interval with $m_h(p_h) = 1$ indicating a perfect match. From this it becomes apparent that the sole purpose of the denominator $(p_h - p_l)$ is to normalize $m_h(p)$. (ii) As long as $p < p_h$, any increase in p improves the quality of the match as p approaches p_h from below. Regarding l-types, their (quality of the) match is $m_l(p) = \frac{p_h - p}{p_h - p_l}$: the smaller p becomes, the larger the value of $m_l(p)$, i.e., the more p matches l's optimal pace p_l.

Under these functional form choices for $m_h(p)$ and $m_l(p)$, teacher utility eq. (5) is maximized at

$$p = n\, p_h + (1 - n)\, p_l =: p^* \tag{6}$$

which becomes quickly apparent from deriving the log of eq. (5), and solving the FOC for p. Eq. (6) states that the optimal pace p^* that maximizes teacher utility is simply a convex combination of each student type's optimal pace. If $n = 0$ ($n = 1$), teachers will choose $p = p_l$ ($p = p_h$). For $0 < n < 1$, however, teachers weight each type's optimal pace by her share.[8]

2.3   Tracked and Mixed School Systems

Once the teacher's problem is solved, one can proceed with modeling the two school systems (tracked and mixed). In tracked systems, students are segregated by type, i.e., h-types (l-types) are enrolled in an upper (lower) level track. Segregation by ability allows teachers to perfectly tailor their instructional pace to each type's optimal pace, however, spillovers are absent. In mixed (or comprehensive) school systems, h-types and l-types are classmates which generates gains from spillovers, but teachers are now forced to set a pace that lies between each type's optimal pace. With a_θ^{tr} (a_θ^{mix}) denoting final achievement of type θ in a tracked (mixed) school system, these notions translate into

$$a_h^{\mathrm{tr}} \equiv a_h(p^*|_{n=1}) = h \qquad\qquad a_h^{\mathrm{mix}} \equiv a_h(p^*|_{n\in(0,1)}) = p^* \tag{7}$$

$$a_l^{\mathrm{tr}} \equiv a_l(p^*|_{n=0}) = l \qquad\qquad a_l^{\mathrm{mix}} \equiv a_l(p^*|_{n\in(0,1)}) = 2p_l - p^*. \tag{7}$$

The four functions are plotted in Figure 1. School systems are represented in the model by the values of n at which the teacher's optimal pace p^* is evaluated. In tracked systems, there exist only classes with $n \in \{0, 1\}$ which simplifies eq. (3) to $a_\theta^{\mathrm{tr}} = \theta$. Comprehensive (or mixed) systems are characterized by $0 < n < 1$. Both a_h^{mix} and a_l^{mix} result from evaluating eq. (3) at p^* while keeping in mind that $p_l < p^* < p_h$ in mixed systems.
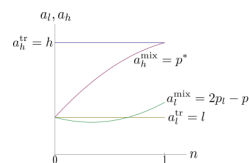
Figure 1: Final achievement of h- and l-types in tracked and mixed school systems. a_l^{mix} and a_h^{mix} are functions of n, see eq. (7). In tracked systems, $a_\theta^{\mathrm{tr}} = \theta$. The values of the exogenous parameters i and $h > l$ are held fixed in this figure.

3   Comparative Statics

Each of the following three subsections establishes one main finding. I first investigate the marginal impact of better peers on each type's final achievement. I then proceed with the effect of "de-tracking", i.e., the shift from a tracked school system towards a comprehensive one. The third subsection derives the condition under which average achievement of all students is maximized under mixing.

3.1   The Effect of Better Peers on Final Achievement

This subsection investigates how changes in the share of high-achievers n affect each type's final achievement $a_\theta(p)$. As this implies $n \in (0,1)$ rather than $n \in \{0,1\}$, the analysis here is entirely based on $a_h^{\text{mix}} = p^*$ and $a_l^{\text{mix}} = 2p_l - p^*$ from eq. (7).

Let's first consider the marginal effect of n on h-types. Because (i) $p_h > p_l$ as long as $n < 1$ or $i < 1$, and (ii) $p^* = n\,p_h + (1-n)\,p_l$, the marginal effect $\frac{\partial a_h^{\text{mix}}}{\partial n} = \frac{\partial p^*}{\partial n}$ must be positive. $\frac{\partial a_h^{\text{mix}}}{\partial n} = \frac{\partial p^*}{\partial n} > 0$ means that increases in n induce teachers to set a more demanding pace, which raises h-types' final achievement.

For l-types, the marginal impact of better peers is the sum of two effects:

$$\frac{\partial a_l^{\text{mix}}}{\partial n} = \underbrace{-\frac{\partial p^*}{\partial n}}_{1.} + \underbrace{2\frac{\partial p_l}{\partial n}}_{2.}. \qquad [8]$$

The first effect, $-\frac{\partial p^*}{\partial n} < 0$, is called (negative) *pace effect*. Contrary to h-types, the pace effect $\frac{\partial p^*}{\partial n}$ has a negative impact on l-types' achievement. This makes sense because l-types – who were already struggling with the current pace (as $p^* > p_l$ whenever $n > 0$) – now face even greater difficulties in keeping up. The second effect, $2\frac{\partial p_l}{\partial n} = 2\frac{\partial s_l}{\partial n} = 2i(h-l) > 0$, called *spillover effect*, is positive because increases in n raise the extent of spillovers s_l which are beneficial for low-achievers. Taken together, marginal increases in n raise a_l^{mix} only if the negative pace effect is overcompensated by the positive spillover effect. Therefore,

Proposition 1: *Let $n, i \in (0,1)$. Then, increases in the share of better students are beneficial for l-types only in classes (n,i) where*

$$i > \frac{1}{1+2n} =: i_1^*$$

and detrimental otherwise. h-types always benefit from better peers because the pace effect is positive for them.

This can be shown easily. First note that the partial derivative of a_l^{mix} from eq. (7) w.r.t. n, i.e. $\frac{\partial a_l^{\text{mix}}}{\partial n} = (h-l)(-1+i+2in)$, is negative for small n. The explanation for this is captured by (8): $\frac{\partial a_l^{\text{mix}}}{\partial n} < 0$ for small values of n as the (positive) spillover effect is dominated by the (negative) pace effect. However, because $\frac{\partial^2 a_l^{\text{mix}}}{\partial n^2} = 2i(h-l) > 0$ for any n, one can further infer that the relative impact of spillover effects must be increasing in n. Therefore, solving $\frac{\partial a_l^{\text{mix}}}{\partial n} = 0$ for i yields the combination of n and i at which the negative pace effect is fully offset by the positive spillover effect. The positive impact of better peers on h-type's achievement follows from the fact that $a_h^{\text{mix}} = p^*$ and $p_h > p_l$ (as long as $n, i < 1$), which implies that $p^* = n\,p_h + (1-n)\,p_l$ must become larger as n increases.

Proposition 1 suggests that marginal increases in n are beneficial for lower-achieving students only in classes (n,i) where $i > i_1^*$. For h-types, however, the pace effect is positive. Therefore, even though h-types do not benefit from spillovers, increases in the share of better students always positively affect their achievement.

So far the analysis shows that marginal increases in n shape each type's final achievement in very different ways. As mentioned in 1, theoretical (empirical) investigations of peer effects usually assume (estimate) models where own and peer achievement are positively related. The parameter on peer achievement is then interpreted as "the" peer effect. As shown here, changes in peer achievement levels – or, equivalently, in n – affect each type's final achievement in different ways, suggesting that the term "peer effect" should be used with caution.

3.2   How Does De-tracking Affect Each Type's Achievement?

A closer inspection of a_l^{mix} reveals that (i) the pace is too challenging for l-types under mixing because – whenever $n < 1$ or $i < 1$ – the teacher's pace must exceed l's optimal pace, i.e., $p^* > p_l$. This implies $a_l^{\text{mix}} = 2p_l - p^* < p_l$: an l-type's final achievement therefore lies below her potential under mixing. (ii) However, as stated in the following proposition, this does not necessarily imply that l-types are hurt from mixing:

Proposition 2: *Let $n, i \in (0,1)$. Then, l-types are better off in mixed systems, i.e. $a_l^{mix} > a_l^{tr}$, if*

$$i > \frac{1}{1+n} =: i_2^*.$$

h-types always prefer tracked systems because $a_h^{\text{tr}} = h > a_h^{\text{mix}} = p^*$.

Figure 3: Summary of findings. $i_2^* = \frac{1}{1+n}$, $i_3^* = \frac{2}{1+2n}$. For classes (n, i) that belong to category I, de-tracking raises both final achievement of low-achievers and average achievement. De-tracking is still beneficial for l-types in category II, but average achievement is depressed. Students enrolled in classes (n, i) that fall into category III should be segregated by ability.

4 Sensitivity Analyzes

This section investigates whether the main result of this study – Proposition 3 – is sensitive to functional form changes of key parts of the model: spillovers s_θ, teacher utility $u(p)$, and final achievement $a_\theta(p)$. The findings presented here are obtained from numerical simulations as the model becomes analytically intractable for alternative functional form choices.[11] As these analyzes compare average achievement under tracking and mixing, the domain of n is restricted to $(0, 1)$ in the following.

The first set of sensitivity checks considers the generalized spillover function

$$s_\theta = (n\,i)^\beta (h - \theta)$$

with $\beta \in \mathbb{R}$. The simulations show that Proposition 3 only holds if $\beta \geq 0.6$. For smaller β, it is possible that $\bar{a}^{\mathrm{mix}} > \bar{a}^{\mathrm{tr}}$ for any $n \in (0, 1)$, i.e., average achievement is maximized under mixing regardless of n. The reason why Proposition 3 is violated in this case is simple: $(n\,i)^\beta > (n\,i)$ if $\beta < 1$, i.e, spillovers become more and more amplified as β decreases.

One should therefore be more concerned about the possibility of an upper bound of β at which spillovers become so small that $\bar{a}^{\mathrm{mix}} < \bar{a}^{\mathrm{tr}}$ for all n. The simulations show that the threshold n^* at which $\bar{a}^{\mathrm{mix}} > \bar{a}^{\mathrm{tr}}$ (for any $n > n^*$) quickly approaches unity as β increases. However, $n^* < 1$ even for very large values of β (e.g., $\beta = 100$). One can therefore conclude that Proposition 3 is valid as long as $\beta \geq 0.6$.

Turning to the teacher's utility function, one may have considered

$$u(p) = a_h(p)^n a_l(p)^{1-n}$$

as a plausible representation of teacher preferences. This function is maximized at

$$p^* = 2n\,p_l = \begin{cases} p_l & \text{for } n \leq 0.5 \\ \min(2n\,p_l, \ h) & \text{for } n > 0.5 \end{cases}.$$

As discussed in 2.2, p^* must lie between p_l and $p_h = h$, implying that teachers choose $p^* = p_l$ if $n \leq 0.5$. Therefore, as long as $n \leq 0.5$, l-types benefit from both the spillovers generated by their abler peers and the optimally targeted pace $p^* = p_l$. The good news is that Proposition 3 turns out to hold, even though the teacher's pacing now heavily favors l-types. Therefore the findings of this study are qualitatively the same, regardless whether $u(p) = m_h(p)^n m_l(p)^{1-n}$ or $u(p) = a_h(p)^n a_l(p)^{1-n}$ is assumed.

Finally, one may wonder to which extent Proposition 3 is sensitive to alternative specifications of the achievement function. Its generalized form

$$a_\theta(p) = p_\theta - (|p - p_\theta|)^\gamma$$

allows for altering the magnitude of the "penalty" for p deviating from p_θ. So far γ was set to unity. The following results are based on the restriction $|p - p_\theta| < 1$, implying that the detrimental impact of $(|p - p_\theta|)^\gamma$ on final achievement becomes larger as γ decreases. The simulations show that $\bar{a}^{\mathrm{mix}} < \bar{a}^{\mathrm{tr}}$ for all n if $\gamma \leq 0.5$, therefore violating Proposition 3. In addition, the graph of $\bar{a}^{\mathrm{mix}}$ looses its S-shape if $\gamma > 3$. Thus Proposition 3 only holds for $\gamma \in (0.5, 3)$.

5 Summary and Conclusions

This paper investigates the impact of de-tracking on achievement levels of low-achievers, high-achievers, and the entire student body by means of a small theoretical model. Final achievement is modeled as a function of a student's type (low or high), knowledge-spillovers, and the instructional pace, which is tailored to the skill distribution of a class.

Four findings emerge. First, student types respond in different ways to changes in the skill composition of a class. Better peers lead to an increase in the instructional pace (pace effect), which is beneficial for good students. Weak students, however, are struggling with the more demanding pace but are also exposed to additional positive knowledge-externalities (spillover effect). Therefore, as the impact of better peers crucially

depends on a student's type, the term "peer effect" should be used with caution. Second, good students are better off in ability-tracked than in mixed school systems, which is in line with most of the empirical evidence. Third, lower-achieving students learn more under tracking if both the share of good students and the extent of interaction between student types are small. This may explain why some (quasi-)experimental studies find negative de-tracking effects for any student type. Fourth, mixing can maximize average achievement if both the share of good students and extent of interaction are high.

This paper highlights that transmission mechanisms and behavioral adjustments of decision makers should be taken into account when evaluating the expected impact of changes in the school system. Carrell, Sacerdote, and West (2013), for example, use reduced form estimates of ability peer effects from a quasi-experimental setting to design student grouping policies that are aimed at helping weaker students. Follow-up assessments, however, reveal that – compared to non-treated weaker students – targeted students were actually hurt by their intervention. The authors therefore conclude that "[the use of] reduced-form estimates to make out-of-sample policy predictions can lead to unanticipated outcomes".[12] The model suggests that comprehensive school systems become more attractive if both the share of high-achievers and extent of interaction are increased. However, policymakers who sympathize with mixed systems should be aware that parents of higher-achieving students may enroll their children in private schools to prevent them from being exposed to weaker public school classmates (Dearden, Ryan, and Sibieta 2011). From the model's view, this problem could be alleviated to some extent by the provision of norms and learning environments that motivate student types interact more.[13] To make school system choices more informed, further research could additionally investigate whether a country's economic prosperity is primarily determined by the average achievement level of its population, or the abilities of (a small number of) exceptionally talented individuals.[14]

Notes

[1]In tracked school systems, students are segregated by academic achievement into different school types or separate classrooms within schools. By contrast, de-tracked/mixed/comprehensive systems are characterized by greater within-class-heterogeneity in skills. Throughout, the terms ability, achievement and skills are used interchangeably.

[2]There are also numerous non-experimental studies on this topic. Based on DiD-estimates, Hanushek and Wößmann (2006) find that early tracking increases educational inequality and may reduce average achievement. Galindo-Rueda and Vignoles (2007) arrive at similar results. Additional studies are summarized and discussed in the comprehensive surveys conducted by Meier and Gabriela (2008) and Betts (2011) .

[3]Throughout this paper, students are female and teachers are male. The gender of students and teachers was decided by coin toss.

[4]Such "monotonous" or "one-dimensional" peer effects are employed in both empirical investigations of ability peer effects (see Ding and Lehrer 2007; Carrell, Fullerton, and West 2009, or Imberman, Kugler, and Sacerdote 2012) as well as theoretical papers on, for example, effort formation (Foster and Frijters 2009), school competition (Fraja and Landeras 2006) or residential segregation (Epple and Romano 1998). The only exception is Lazear (2001). His model differentiates students by their likelihood of exerting disruptive behavior rather than scholastic achievement. However, Lazear (2001) further notes that a student's achievement and disruptive behavior are often negatively correlated which has been shown empirically by Lavy, Paserman, and Schlosser (2011) .

[5]As will be shown later, better peers also have an indirect positive effect on h-types' achievement.

[6]Weinberg (2007) shows empirically that students have stronger social ties with peers who are "similar" to them (known as "homophily" in the sociological literature). Halliday and Kwak (2012) further suggest that mean achievement levels of a student's reference group matter most for her own achievement.

[7]Alternatively, one might hold the curriculum fixed and think of p as the "depth" of coverage in the sense of Carrell and West (2010) .

[8]Alternatively, one could have based teacher utility directly on student achievement $a_\theta(p)$ and suggested $u(p) = a_h(p)^n \cdot a_l(p)^{1-n}$ to represent teacher preferences. As will be shown in 4, however, the results turn out to be qualitatively the same, regardless whether $u(p) = m_h(p)^n m_l(p)^{1-n}$ or $u(p) = a_h(p)^n a_l(p)^{1-n}$ is employed.

[9]For illustrative purposes, one can further compare the graphs of a_l^{tr} and a_l^{mix} in Figure 1.

[10]One may wonder to which extent these findings carried over to settings with, say, three student types l (low), m (middle), and h (high) where h-types' spillovers only increase the potential of m-types, and m-types only affect l-types' learning potential. The effect of an increase in n, the share of h-types, would still be similar to the two-student-type case: an increase in n would indirectly increase an l-type's potential through its direct impact on the potential of m-types. However, as l-types are fully affected by the pace-effect but only receive (potentially weak) indirect spillovers, one should expect that the conditions under which Propositions 1 through 3 hold become more restrictive.

[11]Simulation results are based on the following domains in the exogenous variables: $h \in (0, 2]$, $l \in [0, h)$, and $i \in [0, 1]$. A summary of the sensitivity analyzes is available on request.

[12]This quote is taken from the abstract of the working-paper version Carrell, Sacerdote, and West (2011). In the context of educational production, careful empirical investigations of transmission channels are becoming more common. See, for example, Fraja, Oliveira, and Zanchi (2010) and Lavy and Schlosser (2011), or Pop-Eleches and Urquiola (2013). To empirically test the model, one could operationalize n by grouping students by proficiency levels (see OECD 2011 : 453 as an example) into better and weaker ones. The role of i could be assessed by assuming that within a class, interaction levels among students of the same gender are higher than cross-gender interaction levels (see Crosnoe et al. 2008 and Hill 2015).

[13]Estevan (2016), for instance, finds that higher public school expenditures are associated with reductions in private school enrollment, partly because public schools attract better teachers (Behrman et al. 2016).

[14]The importance of top-achievers for a country's technological and scientific progress is investigated by, among others, Squicciarini and Nico (2015) and Ellison and Swanson (2016). On the other hand, cross-country comparisons conducted by Hanushek and Kimko (2000) and Jamison, Jamison, and Hanushek (2007) show that a one standard deviation increase in *average* math test scores can boost annual GDP growth by up to 1.0pp.

References

Arcidiacono, Peter, Gigi Foster, Natalie Goodpaster, and Josh Kinsler. 2012. "Estimating Spillovers Using Panel Data, with an Application to the Classroom." *Quantitative Economics* 3 (3): 421–470.

Arnott, Richard, and John Rowse. 1987. "Peer Group Effects and Educational Attainment." *Journal of Public Economics* 32 (3): 287–305.

Behrman, Jere, Michela Tincani, Petra Todd, and Kenneth Wolpin. 2016. "Teacher Quality in Public and Private Schools Under a Voucher System: The Case of Chile." *Journal of Labor Economics* 34 (2): 319–362.

Betts, Julian. 2011. "The Economics of Tracking in Education.". In Hanushek, Eric, Stephen Machin, and Ludger Wößmann (Eds.), *Handbook of the Economics of Education*. Vol. 3, 341–381. Amsterdam: Elsevier.

Bursztyn, Leonardo, and Robert Jensen. 2015. "How Does Peer Pressure Affect Educational Investments?" *Quarterly Journal of Economics* 130 (3): 1329–1367.

Card, David, and Laura Giuliano. 2016. "Can Tracking Raise the Test Scores of High-Ability Minority Students?" *American Economic Review* 106 (10): 2783–2816.

Carrell, Scott, and James West. 2010. "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors." *Journal of Political Economy* 118 (3): 409–432.

Carrell, Scott, Bruce Sacerdote, and James West. 2011. "From natural Variation to Optimal Policy? The Lucas Critique Meets Peer Effects." Working Paper 16865, NBER, Cambridge.

Carrell, Scott, Bruce Sacerdote, and James West. 2013. "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation." *Econometrica* 81 (3): 855–882.

Carrell, Scott, Richard Fullerton, and James West. 2009. "Does Your Cohort Matter? Measuring Peer Effects in College Achievement." *Journal of Labor Economics* 27 (3): 439–464.

Crosnoe, Robert, Catherine Riegle-Crumb, Sam Field, Kenneth Frank, and Chandra Muller. 2008. "Peer Group Contexts of Girls' and Boys' Academic Experiences." *Child Development* 79 (1): 139–155.

Dearden, Lorraine, Chris Ryan, and Luke Sibieta. 2011. "What Determines Private School Choice? A Comparison Between the United Kingdom and Australia." *Australian Economic Review* 44 (3): 308–320.

Ding, Weili, and Steven Lehrer. 2007. "Do Peers Affect Student Achievement in China's Secondary Schools?" *Review of Economics and Statistics* 89 (2): 300–312.

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 (5): 1739–1774.

Eisenkopf, Gerald. 2010. "Peer Effects, Motivation, and Learning." *Economics of Education Review* 29 (3): 364–374.

Eisenkopf, Gerald, Zohal Hessami, Urs Fischbacher, and Heinrich Ursprung. 2015. "Academic Performance and Single-Sex Schooling: Evidence from a Natural Experiment in Switzerland." *Journal of Economic Behavior and Organization* 115: 123–143.

Ellison, Glenn, and Ashley Swanson. 2016. "Do Schools Matter for High Math Achievement? Evidence from the American Mathematics Competitions." *American Economic Review* 106 (6): 1244–1277.

Epple, Dennis, and Richard Romano. 1998. "Competition Between Private and Public Schools, Vouchers, and Peer-Group Effects." *American Economic Review* 88 (1): 33–62.

Estevan, Fernanda. 2016. "Public education Expenditures and Private School Enrollment." *Canadian Journal of Economics* 48 (2): 561–584.

Foster, Gigi, and Paul Frijters. 2009. "Are There Complementarities in Educational Peer Effects?" Working Paper.

Foster, Gigi, and Paul Frijters. 2010. "Students' Beliefs About Peer Effects." *Economics Letters* 108 (3): 260–263.

Fraja, Gianni, and Pedro Landeras. 2006. "Could do Better: The Effectiveness of Incentives and Competition in Schools." *Journal of Public Economics* 90 (1-2): 189–213.

Fraja, Gianni, Tania Oliveira, and Luisa Zanchi. 2010. "Must try Harder: Evaluating the Role of Effort in Educational Attainment." *Review of Economics and Statistics* 92 (3): 577–597.

Fruehwirth, Jane. 2013. "Identifying Peer Achievement Spillovers: Implications for Desegregation and the Achievement Gap." *Quantitative Economics* 4 (1): 85–124.

Galindo-Rueda, Fernando, and Anna Vignoles. 2007. "The Heterogenous Effect of Selection in UK Secondary Schools.". In Wößmann, Ludger, and Paul Peterson (Eds.), *Schools and the Equal Opportunity Problem, Chapter 5*. 103–128. Cambridge: MIT Press.

Halliday, Timothy, and Sally Kwak. 2012. "What is a Peer? The Role of Network Definitions in Estimation of Endogenous Peer Effects." *Applied Economics* 44 (3): 289–302.

Hanushek, Eric, and Dennis Kimko. 2000. "Schooling, Labor-Force Quality and the Growth of Nations." *American Economic Review* 90 (5): 1184–1208.

Hanushek, Eric, and Ludger Wößmann. 2006. "Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence Across Countries." *Economic Journal* 116 (510): 63–76.

Hidalgo-Hidalgo, Marisa. 2011. "On the Optimal Allocation of Students When Peer Effects Are at Work: Tracking vs. Mixing." *SERIEs: Journal of the Spanish Economic Association* 2 (1): 31–52.

Hill, Andrew. 2015. "The Girl Next Door: The Effect of Opposite Gender Friends on High School Achievement." *American Economic Journal: Applied Economics* 7 (3): 147–177.

Hoxby, Caroline. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." Working Paper 7867, NBER, Cambridge.

Imberman, Scott, Adriana Kugler, and Bruce Sacerdote. 2012. "Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees." *American Economic Review* 102 (5): 2048–2082.

Jain, Tarun, and Mudit Kapoor. 2015. "The impact of Study Groups and Roommates on Academic Performance." *Review of Economics and Statistics* 97 (1): 44–54.

Jamison, Eliot, Dean Jamison, and Eric Hanushek. 2007. "The Effects of Education Quality on Income Growth and Mortality Decline." *Economics of Education Review* 26 (6): 771–788.

Lavy, Victor, and Analia Schlosser. 2011. "Mechanisms and Impacts of Gender Peer Effects at School." *American Economic Journal: Applied Economics* 3 (2): 1–33.

Lavy, Victor, Daniele Paserman, and Analia Schlosser. 2011. "Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom." *Economic Journal* 122 (559): 208–237.

Lazear, Edward. 2001. "Educational Production." *Quarterly Journal of Economics* 116 (3): 777–803.

Lovell, John. 1960. "The Bay High School Experiment." *Educational Leadership* 17: 383–387.

Marascuilo, Leonard, and McSweeney. Maryellen. 1972. "Tracking and Minority Student Attitudes and Performance." *Urban Education* 6 (4): 303–319.

Meier, Volker. 2004. "Choosing Between School Systems: The Risk of Failure." *FinanzArchiv* 60 (1): 83–93.

Meier, Volker, and Schütz. Gabriela. 2008. "The Economics of Tracking and Non-Tracking." *Zeitschrift für Betriebswirtschaft* Special Issue 1: 23–43.

OECD. 2011. *Education at a Glance 2011: OECD Indicators* Paris: OECD Publishing.

Pekkarinen, Tuomas, Roope Uusitalo, and Sari Kerr. 2013. "School Tracking and Development of Cognitive Skills." *Journal of Labor Economics* 3 (3): 577–602.

Pop-Eleches, Cristian, and Miguel Urquiola. 2013. "Going to a Better School: Effects and Behavioral Responses." *American Economic Review* 103 (4): 1289–1324.

Squicciarini, Mara, and Voigtländer. Nico. 2015. "Human Capital and Industrialization: Evidence from the Age of Enlightenment." *Quarterly Journal of Economics* 130 (4): 1825–1883.

Weinberg, Bruce. 2007. "Social Interactions with Endogenous Associations." Working Paper 13038, NBER, Cambridge.

Whitmore, Diane. 2005. "Resource and Peer Impacts on Girls' Academic Achievement: Evidence from a Randomized Experiment." *American Economic Review* 95 (2): 199–203.