A model about the impact of ability grouping on student achievement

David Kiss*

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.

JEL: I24, J24, H52

Keywords: tracking vs. mixing, decomposition of ability peer effects.

*Leibniz University Hannover, Institute of Labor Economics, Koenigsworther Platz 1, 30167 Hannover, Germany, ✉: kiss@aoek.uni-hannover.de, ☎: +49 511 762 5620. A revised version of this paper was published under the same title in 2017 in *The B.E. Journal of Economic Analysis & Policy*, 17(3), doi.org/10.1515/bejeap-2017-0036. I want to thank Pavel Jelnov, Patrick Puhani, Regina Riphahn, Richard Romano, and Guido Schwerdt for very valuable comments. This paper benefited a lot from the remarks and suggestions made by two anonymous referees. Financial support by the Bavarian Graduate Program in Economics (BGPE) is gratefully acknowledged. Any errors or omissions are mine.

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 et al. 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 et al. (2013) find that de-tracking raises average achievement in their sample of Finish secondary students.

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

[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.

effect on weaker students' learning (spillover effect).

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.

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 summarized in Table 1, also 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.[2]

The remainder of the paper is as follows. The model is derived in Section 2. Comparative statics analyzes are presented in Section 3. Section 4 concludes.

[2] Such "monotonous" or "one-dimensional" peer effects are employed in both empirical investigations of ability peer effects (see Ding and Lehrer 2007, Carrell et al. 2009, or Imberman et al. 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). Throughout this paper, students are female and teachers are male. The gender of students and teachers was decided by coin toss.

2 The Model

2.1 Determinants of final achievement

There are two student types

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

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

$$p_\theta \equiv \theta + s_\theta \tag{1}$$

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 et al. 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, Bursztyn and Jensen 2015). Generally speaking, h-types may serve as role models.[3]

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)

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

may easily establish friendships with their higher-achieving male (female) peers, which would be reflected by a high value of i (Whitmore 2005, 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, Fruehwirth 2013).[4]

Evaluation of (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.[5] 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

[4] 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.

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

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 sub-populations 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 (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 type's 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 (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 (5) is maximized at

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

which becomes quickly apparent from deriving the log of (5), and solving the FOC for

p. (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.[6]

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

$$\begin{aligned} a_h^{\text{tr}} &\equiv a_h(p^*|_{n=1}) = h \qquad & a_h^{\text{mix}} &\equiv a_h(p^*|_{n\in(0,1)}) = p^* \\ a_l^{\text{tr}} &\equiv a_l(p^*|_{n=0}) = l \qquad & a_l^{\text{mix}} &\equiv a_l(p^*|_{n\in(0,1)}) = 2p_l - p^*. \end{aligned} \tag{7}$$

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 (3) to $a_\theta^{\text{tr}} = \theta$. Comprehensive (or mixed) systems are characterized by $0 < n < 1$. Both a_h^{mix} and a_l^{mix} result from evaluating (3) at p^* while keeping in mind that $p_l < p^* < p_h$ in mixed systems.

[6] 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. However, that utility function is maximized at $p^* = 2n\,p_l$. Therefore, as long as $n \leq 0.5$, teachers will choose $p^* = p_l$. This means that teachers will perfectly match l-types' optimal pace as long as n is smaller than 50%, which is implausible. Preference is further given to $u(p) = m_h(p)^n \cdot m_l(p)^{1-n}$ to (greatly) increase the analytical tractability of the model. Regarding the main findings presented in Section 3, however, numerical simulations under $u(p) = a_h(p)^n \cdot a_l(p)^{1-n}$ turn out to produce very similar results. That is, results 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 employed.

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 (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:

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

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 (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 Section 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 $a_h^{tr} = h > a_h^{mix} = p^*$.

First note that the precondition $n, i \in (0,1)$ implies $\frac{\partial a_l^{\text{mix}}}{\partial n} < 0$ for small n, and $\frac{\partial^2 a_l^{\text{mix}}}{\partial n^2} > 0$ (for any n) for the same reasons given in Proposition 1. Hence, for $i = i_2^*$, the positive sign of $\frac{\partial^2 a_l^{\text{mix}}}{\partial n^2}$ means that spillover effects might be strong enough to overcompensate both the negative pace effect and the mismatch between current and l's optimal pace.[7] i_2^* is obtained by solving $a_l^{\text{mix}} = a_l^{\text{tr}}$ for i. The statement concerning h-types can be directly inferred from (7) by comparing a_h^{tr} with a_h^{mix} because $p^* < p_h = h$ for $n, i < 1$.

Proposition 2 suggests that mixed systems are preferred by l-types in classes (n, i) where both the extent of interaction and the share of h-types are sufficiently high. The required level of interaction at which l-types are indifferent between tracking and mixing is decreasing in n because $\frac{\partial i_2^*}{\partial n} < 0$. h-types are hurt under mixing as their weaker classmates induce teachers to set a pace that is below their learning potential $p_h = h$.

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

3.3 The impact of de-tracking on average achievement

The last policy-relevant variable investigated here is the average achievement level of all students

$$\bar{a} \equiv n\, a_h(p^*) + (1-n)\, a_l(p^*),$$

which is a weighted average of each type's final achievement under p^*. In tracked systems, i.e. for $n \in \{0, 1\}$ within each single class, average achievement of *all* students equals $\bar{a}^{\text{tr}} \equiv n\, h + (1-n)l$. To be more elaborate: if the size of the entire student body is normalized to unity, then the *share* of classes composed solely of h-types and l-types becomes n and $(n-1)$, respectively. Regarding mixed school systems, one can interpret $\bar{a}^{\text{mix}} \equiv n\, a_h^{\text{mix}} + (1-n)a_l^{\text{mix}}$ as the final achievement level of a single representative class. Comparability between $\bar{a}^{\text{tr}}$ and $\bar{a}^{\text{mix}}$ is therefore established by normalizing of the size of the (entire) student body to unity.

Both $\bar{a}^{\text{tr}}$ and $\bar{a}^{\text{mix}}$ are plotted in Figure 2. The graph of $\bar{a}^{\text{mix}}$ turns out to be S-shaped: for n smaller than some threshold n^*, average achievement in mixed systems is smaller than in tracked ones. However, for $n \geq n^*$, mixed systems yield higher average achievement levels, which is formalized in the following proposition:

Proposition 3. *Let $n, i \in (0, 1)$. Average achievement is greater in mixed systems if $\bar{a}^{mix} > \bar{a}^{tr}$ or, equivalently,*

$$i > \frac{2}{1+2n} =: i_3^*.$$

As shown in Proposition 1, $\frac{\partial a_l^{\text{mix}}}{\partial n} < 0$ for small n. At the same time, $\frac{\partial \bar{a}^{\text{tr}}}{\partial n} > 0$ for all n because $h > l$. Therefore, both $\lim_{n\to 0} \frac{\partial \bar{a}^{\text{mix}}}{\partial n} = \lim_{n\to 0} \frac{\partial a_l^{\text{mix}}}{\partial n} < 0 < \left.\frac{\partial \bar{a}^{\text{tr}}}{\partial n}\right|_{n=0}$ and $\lim_{n\to 0} \bar{a}^{\text{mix}} = \left.\bar{a}^{\text{tr}}\right|_{n=0}$ imply $\bar{a}^{\text{tr}} > \bar{a}^{\text{mix}}$ for small values of n. Consequently, if $\bar{a}^{\text{mix}} = \bar{a}^{\text{tr}}$ can be solved for i (which will denote i_3^*), there must be a value n^* at which both $\left.\bar{a}^{\text{mix}}\right|_{n^*} = \left.\bar{a}^{\text{tr}}\right|_{n^*}$ and $\left.\frac{\partial \bar{a}^{\text{mix}}}{\partial n}\right|_{n^*} > \left.\frac{\partial \bar{a}^{\text{tr}}}{\partial n}\right|_{n^*}$ must hold. Because of that, it must be the case that $\bar{a}^{\text{mix}}\Big|_n > \bar{a}^{\text{tr}}\Big|_n$ for $n \in (n^*, 1)$, with n^* simply denoting the inverse of i_3^*.

Classes (n, i) with $i > i_3^*$ exhibit higher average achievement levels if students are allowed to learn together instead of being segregated by ability. As expected, $i_3^* > i_2^*$ for any n because spillovers on l-types now have to overcompensate the suboptimal pace for both student types.

The three propositions are summarized in Figure 3. From the model's perspective, a class (n, i) can be classified into one of the following three categories. In category I, de-tracking rises both average achievement and l-types' achievement. De-tracking is still beneficial for l-types in category II, but average achievement is depressed. Students should be segregated by ability if classes (n, i) fall into category III.

4 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 consideration when evaluating the expected impact of changes in the school system. Carrell et al. (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".[8]

The model suggests that comprehensive school systems become more attractive if both the share of high-achievers and extent of interaction are increased. Probably the extent of interaction could be raised in less time (say, the medium run) at a lower cost. 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.[9]

---

[8] This quote is taken from the abstract of the working-paper version Carrell et al. (2011). In the context of educational production, careful empirical investigations of transmission channels are becoming more common. See, for example, Fraja et al. (2010), Lavy and Schlosser (2011), or Pop-Eleches and Urquiola (2013).

[9] The importance of top-achievers for a country's technological and scientific progress is investigated by, among others, Squicciarini and Voigtländer (2015) and Ellison and Swanson (2016). On the other hand, cross-country comparisons conducted by Hanushek and Kimko (2000) and Jamison et al. (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.

Betts, Julian. 2011. "The economics of tracking in education." Vol. 3 of *The Handbook of the Economics of Education*, ed. by Eric Hanushek, Stephen Machin and Ludger Wößmann, Chapter 7, 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." NBER Working Paper 16865, 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.

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): 1244–77.

Epple, Dennis, and Richard Romano. 1998. "Competition between private and public schools, vouchers, and peer-group effects." *American Economic Review*, 88(1): 33–62.

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.

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.

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.

Hoxby, Caroline. 2000. "Peer effects in the classroom: Learning from gender and race variation." NBER Working Paper 7867, Cambridge: National Bureau of Economic Research.

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): 771–88.

Lavy, Victor, and Analia Schlosser. 2011. "Mechanisms and impacts of gender peer effects at school." *American Economic Journal: Applied Economics*, 3(2): 1–33.

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

Marascuilo, Leonard, and Maryellen McSweeney. 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.

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.

Slavin, Robert. 1990. "Achievement effects of ability grouping in secondary schools: A best-evidence synthesis." *Review of Educational Research*, 60(3): 471.

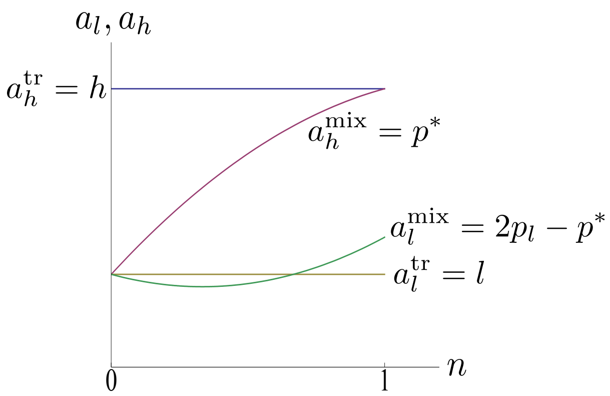
Squicciarini, Mara, and Nico Voigtländer. 2015. "Human capital and industrialization: Evidence from the Age of Enlightenment." *Quarterly Journal of Economics*, 130(4): 1825–1883.

Weinberg, Bruce. 2007. "Social interactions with endogenous associations." NBER Working Paper 13038, Cambridge: National Bureau of Economic Research.

Whitmore, Diane. 2005. "Resource and peer impacts on girls' academic achievement: Evidence from a randomized experiment." *American Economic Review*, 95(2): 199–203.

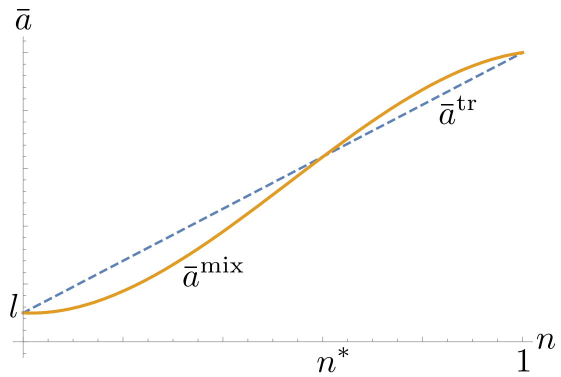
Figures

Figure 1: Final achievement of h- and l-types in tracked and mixed school systems



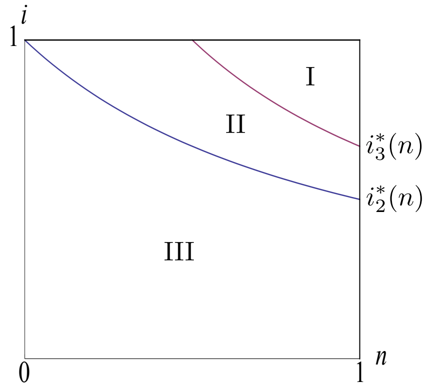
a_l^{mix} and a_h^{mix} are functions of n, see (7). In tracked systems, $a_\theta^{tr} = \theta$. The values of the exogenous parameters i and $h > l$ are held fixed in this figure.

Figure 2: Average achievement levels in mixed and tracked school systems



$\bar{a}^{mix}$ ($\bar{a}^{tr}$) is represented by the solid (dashed) line. One can observe that $\bar{a}^{mix} \geq \bar{a}^{tr}$ only for n greater than n^*.

Figure 3: Summary of findings



$$\begin{array}{rlcl}
\text{I:} & \bar{a}^{\text{mix}} > \bar{a}^{\text{tr}} & \text{and} & a_l^{\text{mix}} > a_l^{\text{tr}} \\
\text{II:} & \bar{a}^{\text{mix}} < \bar{a}^{\text{tr}} & \text{and} & a_l^{\text{mix}} > a_l^{\text{tr}} \\
\text{III:} & \bar{a}^{\text{mix}} < \bar{a}^{\text{tr}} & \text{and} & a_l^{\text{mix}} < a_l^{\text{tr}}
\end{array}$$

$i_2^* = \frac{1}{1+n}$, $i_3^* = \frac{2}{1+2n}$. For classes (n, i) that belong to category I, de-tracking rises both average achievement and final achievement of low-achievers. 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.

Tables

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

Study	Country	Impact of de-tracking on...		
		Low-ach.	High-ach.	Mean ach.
Empirical (tracking experiments)				
Lovell (1960)	USA	$-$	$--^{***}$	$-^{**}$
Duflo et al. (2011)	Kenya	$-^{**}$	$-^{**}$	$-^{**}$
Card and Giuliano (2016)	USA	none	$-^{***}$	$-^{\text{n.a.}}$
Marascuilo and McSweeney (1972)	USA	$++^{**}$	$-$	$+$
Pekkarinen et al. (2013)	Finland	$++^{**}$	$+$	$+^{**}$
Theoretical				
Hidalgo-Hidalgo (2011)		$+$	$-$	$-$
Arnott and Rowse (1987)		$+$	$-$	$\pm$
This paper		$\pm$	$-$	$\pm$
Meier (2004)		$\pm$	$\pm$	$\pm$

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. Betts (2011) and Slavin (1990) provide comprehensive surveys of the empirical tracking literature.
Depending on the values of the exogenous parameters, the theoretical models in the bottom panel can produce a range of de-tracking effects. $\pm$ means that a model can generate either negative, none, or positive de-tracking effects.