CESifo WORKING PAPERS

7866
2019

September 2019

Academic Achievement and Tracking – A Theory Based on Grading Standards

Tim Ehlers, Robert Schwager

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7866*
Category 5: Economics of Education

Academic Achievement and Tracking – A Theory Based on Grading Standards

Abstract

We present a theory explaining the impact of ability tracking on academic performance based on grading policies. Our model distinguishes between initial ability, which is mainly determined by parental background, and eagerness to extend knowledge. We show that achievements of low ability students may be higher in a comprehensive school system, even if there are neither synergy effects nor interdependent preferences among classmates. This arises because the comprehensive school sets a compromise standard which exceeds the standard from the low ability track. Moreover, if students with lower initial ability have higher eagerness to learn, merging classes will increase average performance.

JEL-Codes: I210, I280, D630.

Keywords: ability tracking, comprehensive school, education, equality of opportunity, peer group effects.

Tim Ehlers
Department of Economics
University of Goettingen
Platz der Goettinger Sieben 3
Germany – 37073 Goettingen
tehlers@uni-goettingen.de

*Robert Schwager**
Department of Economics
University of Goettingen
Platz der Goettinger Sieben 3
Germany – 37073 Goettingen
rschwag@uni-goettingen.de

*corresponding author

I Introduction

A major controversy in education policy concerns whether students should be taught in comprehensive schools or whether classes should be tracked according to ability. Proponents of comprehensive schooling argue that mixing good and mediocre students improves the performance of mediocre students without harming the good students too much. This reasoning is largely supported by empirical research, surveyed below, which often finds that tracking, or the ability composition of classes, affect the performance of individual students.

While such peer group effects seem to be well documented empirically, the mechanism driving them is rarely discussed and remains controversial. Explanations typically emphasize the interactions of students. For example, good students might help mediocre colleagues to pass the exam and in addition learn by explaining the subject, or students may follow norms set by other students when choosing their learning effort. While we do not question the relevance of these explanations, in this paper we present a complementary theory which is based on the schools' grading policy. We show that the incentives created by grading standards alone can explain many empirical results on tracking without referring to any direct impact of classmates' ability on individual performance, and without assuming any interdependence of students' preferences.

A large literature has emerged which empirically analyzes peer group effects and the impact of tracking on educational outcomes. Surveys of this literature are provided by Brunello and Checchi (2007) and Meier and Schütz (2008). While peer effects have been shown to arise from various characteristics, including ethnicity (Friesen and Krauth, 2010), gender (Jahanshahi, 2017), body weight (Asirvatham et al., 2018), and age (Foureaux Koppensteiner, 2018), the main focus of research has been on ability. In particular, the impact of tracking on average academic performance and on performance of students with different abilities has attracted much attention. The latter question is linked to equality of opportunity, in the sense that academic achievement should not depend on the social background. This is especially relevant in the case of early tracking, since family background is likely to strongly influence the ability in the first years of schooling.

While some researchers find no evidence that low ability students are harmed in a tracked environment (see Betts and Shkolnik, 2000; Figlio and Page, 2002; Kim et al., 2008), most studies agree that the impact of family background is reinforced by early tracking, and that equality of opportunity is promoted by comprehensive schooling (see Argys et al., 1996; Pekkarinen, 2008; Wößmann, 2010; Ruhose and Schwerdt, 2016). In contrast, there does not seem to be clear evidence on whether average achievement rises or falls when students are tracked (see Hanushek and Wößmann, 2006).

Theoretical models of peer effects usually assume a learning production function where a student's performance does not only depend on his or her ability and effort, but is also positively affected by the average ability in the class the student attends. An early formalization of this kind of externality is provided by Arnott and Rowse (1987). The models by Epple, Newlon and Romano (2002), Duflo, Dupas and Kremer (2011), and Hidalgo-Hidalgo (2014) apply this mechanism to tracking. A second approach to model peer effects is based upon interdependent preferences, where a student's utility depends on characteristics or actions of other students. Such interdependence can arise when students have a preference for ranking high in class (Damiano, Li and Suen, 2010), from network relationships among students (Liu, Patacchini and Zenou, 2014), or from reference point dependent preferences where the reference point is given by other students' performance (Thiemann, 2017).

In this paper we provide a simple model which can account for some of the empirical facts summarized above. In contrast to the theoretical literature cited in the previous paragraph, we do not rely on any externality in the education production function or on any direct effect of other classmates on a student's utility function. We also do not assume that teachers' effort or school resources change when the composition of classes changes, as in Duflo, Dupas and Kremer (2011). Hence, we offer a complementary, and remarkably simple, explanation for important empirical facts.

In our model there are two types of students distinguished by ability. These students are taught either in tracked classes or in a comprehensive school. The instrument of the school is the graduation standard, which is the level of performance required to pass the exam. Like in the classical models of grading by Costrell (1994) and Betts (1998), the school sets the standard by trading off wages of graduates, which rise in the standard, against effort costs required to meet a more demanding standard.

As a major innovation, our model features two different dimensions of ability. The first dimension represents the endowment a student starts with. This endowment is the ability a student has when the tracking decision is taken. It results from previous learning, which is determined by family background and former schooling. The second dimension of ability is the eagerness to learn or to improve personal achievement. This is the student's potential a teacher can work with. This dimension determines how hard it is for students to raise performance beyond their initial ability. Importantly, we do not exclude the case where students with low initial endowment have high eagerness to learn or vice versa.

A further special characteristic of our model is that we distinguish between the objective function of the student or teacher, and the objective parents or society may have. We assume that the teacher, or the school authority which sets grading guidelines for teachers, takes students' effort costs into account when setting the standard. This objective reflects the widespread concern about students feeling excessively pressured by school (see World Health Organization, 2016, p. 59-62). In contrast, in line with the emphasis which policy makers and researchers put on performance measuring tests like PISA and TIMMS, the social objective consists of maximizing performance and disregards effort costs.

In this framework we characterize the standards chosen by tracked schools and by a comprehensive school, and compare the resulting academic achievements of both types of students. We show that students with lower initial endowment gain from a comprehensive school when their eagerness to learn is not too different from the eagerness of students with high initial endowment. This arises because in a comprehensive school the teacher is forced to set the standard as a compromise, which pushes lower ability students to higher achievement.

In a further result we show that the average performance in the comprehensive system is higher if students with lower initial endowment have higher eagerness to learn. In this case the tracked system does not make full use of the learning potential of students with low initial endowment, on which tracking is based. In contrast, if students with low initial endowment also have lower eagerness to learn, average performance goes down when classes are merged.

It is worth noting that these results crucially depend on the two specific elements of our model. First, when all students have the same eagerness to learn and hence differ only in one dimension of ability, then a compromise standard at the comprehensive school necessarily leads to the same average performance as the separate standards of tracked schools. Second, in our model non-tracking will always be dominated by tracking if teachers and society share the same objective function. In the comprehensive school a unique standard must be chosen and one degree of freedom is given up. Therefore maximization of either average performance or welfare is carried out under an additional restriction in the comprehensive school, and a tracked school system is always preferable.

Our paper contributes to the theory of grading, initiated by Costrell (1994) and Betts (1998). This line of research is particularly concerned with the information content of grades and the resulting incentives for schools to award better grades than students deserve, a policy which is often termed 'grade inflation' (Chan, Li and Suen, 2007; Ostrovsky and Schwarz, 2010; Popov and Bernhardt, 2013; Zubrickas, 2015).

Different from the focus of this literature, in this paper, we analyze the impact of grading policies on the achievements of students in tracked and comprehensive classes. This issue is similar to the comparison of a centralized or decentralized standards as in Costrell (1997), since the former is uniform and the latter can be differentiated. In addition Costrell's model, similarly to ours, allows for differences in abilities among schools. However, the focus of his analysis is different from ours. His main issue is that individual schools can free-ride on the tough standards of other schools in the case where employers can only observe the average achievement of graduates from all schools. In contrast we focus on tracking according to ability. Moreover, as stressed above, our model extends Costrell's setup by assuming two dimensions of ability and by replacing the ever positive marginal costs of learning by the idea that a certain performance level can be reached without costs.

The remainder of the paper is organized as follows. In Section II, we present the model and analyze the standards set by tracked schools. Section III contains the analysis of the standard set by a comprehensive school, and in Section IV we compare performance across both systems. The final Section V provides some concluding remarks. Proofs and some formal derivations are relegated to the supplementary material.

II The model with ability tracking

There are two types of students $i \in \{l;h\}$ which differ in ability. We consider two dimensions of ability. The first dimension, which we label endowment or initial ability, plays two roles in the model. First, it is the achievement level which is reached at the time of tracking, and hence serves as the basis of assigning students to schools if tracking occurs. Second, the endowment represents the level of performance a student can achieve without feeling stressed. This is motivated by the idea that students actually like to think, solve problems, and participate in class, at least up to some point, and that they feel bored if courses do not challenge them enough. Each individual of type $i \in \{l;h\}$ has the same initial endowment γ_i. We assume $\gamma_l < \gamma_h$, so that students of type l have lower endowment than students of type h.

The second dimension of ability, labeled a_i for type i, expresses the ease of learning. This parameter measures how much stress a student feels if he or she pushes performance beyond his or her initial ability γ_i. We can interpret a_i as the individual intellectual capacity and motivation of the student.

This modeling captures the fact that both innate ability and parental background are relevant for academic achievement, but differently so at different age levels. We think of early achievement, expressed by γ_i, as being largely determined by the upbringing and the parental background of students, as mentioned in the introduction. This affects the level of performance achievable without feeling stressed, since the latter most likely depends on previous learning. In contrast, we maintain that the ease with which a student moves beyond this level of achievement in later stages of education, expressed by a_i, depends relatively more on individual ability than on background.

Since a student with a low academic background can well be highly motivated or intelligent, we allow for the case $a_l > a_h$. This describes the situation where students of type l have low initial endowment of ability but high eagerness and capacity to learn. Thus, our two-dimensional model of learning captures in a simple way the idea that students who might be assigned to a lower track may have a learning potential superior to the observed performance at the time of tracking.

Depending on both dimensions of ability a student has costs c_i to achieve a certain level of education, denoted by e_i:

$$c_i(e_i) = \frac{1}{2a_i}(e_i - \gamma_i)^2 . \tag{1}$$

Here the inverse of a_i enters the marginal cost of learning. There is a minimum at $e_i = \gamma_i$, where costs are zero. At this point the student's academic performance is just his or her initial ability. The cost function, shown in figure 1, represents the idea that demanding less effort leads to higher costs in terms of being bored.

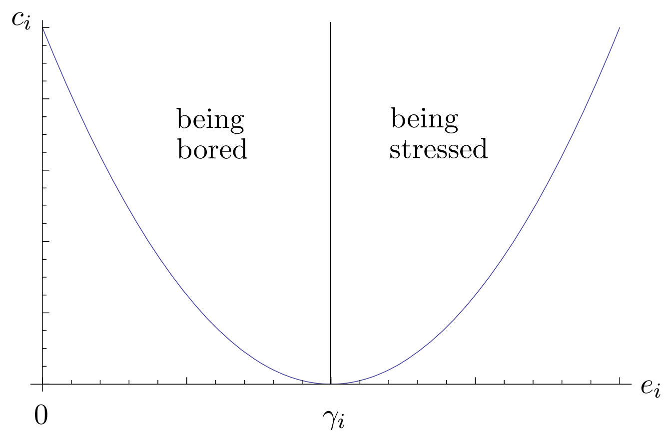


Figure 1: Cost as a function of effort for a student of type i. When effort equals initial ability γ_i, marginal cost of effort is zero.

We now turn to the examination and the labor market. We start in this section by analyzing the case where students are taught in classes tracked according to initial endowment. The school or the teacher set a standard s_i which is measured in the same units as the level of education e_i. The level of personal education must be at least as high as the standard of the school in order to pass the final exam and graduate. The students decide to become graduates or not and, conditional on this, which academic performance to achieve. This decision is based on the cost of effort and the wages for non-graduates w_{0i} and graduates w_{1i}.

The formulation of wages incorporates two assumptions on the information of employers. First, the individual performance e_i is unknown to the employer and therefore does not enter the wage. This assumption is common in theories of grading or examination standards as, for example, Costrell (1994) and Betts (1998). If, instead, the individual performance was known, whether a student meets the standard or not would not convey any additional information, and hence standards would be redundant. Second, employers observe a student's type i and whether he or she graduated or not. This seems straightforward in the tracked schooling system, since employers can derive the type of students from observing the school he or she graduated from. However, in our view the type is not only linked to the school, but is a characteristic of the individual person. This is likely to be observed on the basis of criteria such as address, name, parental occupation, former schooling, behavior, attitude, and manner of speaking, especially when type is linked to social origin. Therefore, it is plausible that a student's type is still observable when he or she attended a comprehensive school.

Conditional on the decision to pass the exam or not, the student chooses performance to maximize utility. We denote the utility achieved in case of passing (not passing) by u_{1i} (u_{0i}):

$$\begin{aligned} u_{1i} &= \max_{e_i}\{w_{1i} - c_i(e_i) | e_i \geq s_i\} \quad \Rightarrow \begin{cases} e_i = s_i & \text{if } s_i \geq \gamma_i \\ e_i = \gamma_i & \text{if } s_i < \gamma_i \end{cases} \\ u_{0i} &= \max_{e_i}\{w_{0i} - c_i(e_i)\} \quad \Rightarrow e_i = \gamma_i\,. \end{aligned} \tag{2}$$

Using this choice, a student graduates if $u_{1i} \geq u_{0i}$. We assume that students expect $w_{1i} \geq w_{0i}$. One can see from equation (4) below that this expectation is confirmed in equilibrium. The utility resulting from the optimal graduation choice is then given by:

$$\max\{u_{0i}; u_{1i}\} = \begin{cases} u_{1i} & \text{if } s_i < \gamma_i \\ u_{1i} & \text{if } w_{1i} - c_i(s_i) - w_{0i} \geq 0 \quad \text{and } s_i \geq \gamma_i \\ u_{0i} & \text{if } w_{1i} - c_i(s_i) - w_{0i} < 0 \quad \text{and } s_i \geq \gamma_i\,. \end{cases} \tag{3}$$

In equilibrium, the wage after passing the exam w_{1i} must be equal to the expected productivity of graduates of class i. We normalize productivity to be measured in the same units as academic performance. Therefore w_{1i} equals the education level of graduates of class i. In the same way the wage w_{0i} is given by the academic performance of non-graduates. From equation (2) we have:

$$\begin{aligned} w_{0i} &= \gamma_i & &\Rightarrow \text{no exam} \\ w_{1i} &= \max\{s_i; \gamma_i\} & &\Rightarrow \text{exam}\,. \end{aligned} \tag{4}$$

We now turn to the choice of standard s_i by the teacher. The teacher maximizes utility of all students. Thus, we assume that the teacher cares about the disutility of learning of his or her students. Inserting (4) and (1) into (3) shows that in the case where $s_i \geq \gamma_i$, the student chooses to graduate if $s_i - \gamma_i \leq 2a_i$. Using this, (2) and (4) in (3) shows that utility of all students of type i is given by:

$$V_i(s_i) = \begin{cases} \gamma_i & \text{if } s_i < \gamma_i \text{ or } s_i - \gamma_i > 2a_i \\ s_i - \frac{1}{2a_i}(s_i - \gamma_i)^2 & \text{if } s_i \geq \gamma_i \text{ and } s_i - \gamma_i \leq 2a_i\,. \end{cases}$$

The optimal standard s_i^* is determined by the first order condition:

$$\frac{\partial V_i}{\partial s_i} = 1 - \frac{1}{a_i}(s_i - \gamma_i) \stackrel{!}{=} 0 \quad \Rightarrow \quad s_i^* = a_i + \gamma_i\,. \tag{5}$$

From strict concavity of the second line in $V_i(s_i)$ and $\gamma_i < s_i^* < \gamma_i + 2a_i$, this is a global optimum. The chosen standard reflects both dimensions of ability. The indirect utilities reached by both types are

$$V_i(s_i^*) = \gamma_i + \frac{a_i}{2} \quad \text{for } i = l, h\,. \tag{6}$$

Comparing the two standards, the typical case is given by $s_h^* > s_l^*$, where h-students enjoy an advantage compared to l-students in terms of total ability:

$$a_h + \gamma_h > a_l + \gamma_l\,. \tag{7}$$

The behavioral rule (2) and the ranking of standards according to (7) show the implications of having two dimensions of ability, linked by the effort cost function (1), in the model. Students with high initial ability γ_h, procured by favorable background, will perform at that level even if standards are low or if they decide not to pass the exam. As is apparent from (2), for this outcome, the precise shape of the cost function for $e_i < \gamma_i$ is irrelevant as long as costs are decreasing. By implication, these students can be held to quite demanding standards even when they are not particularly able to, or interested in, learning more.

In the case of inequality (7) one can clearly label both types of students as l-low and h-high ability. However we do not rule out the opposite case, where

$$a_h + \gamma_h \leq a_l + \gamma_l\,. \tag{8}$$

Thus, we allow the learning capacity of l-students to be so much higher than the one of h-students that it overcompensates the disadvantage of initial endowment of the l-students.

III Merging classes

The previous analysis dealt with separated classes $i \in \{l;h\}$. In contrast in this section we consider the case where both classes can be mixed together in one comprehensive school. We denote the share of h-students in the comprehensive school by $0 < d < 1$. The teacher sets a common standard s applying to all students in the mixed class. The teacher's objective function is the aggregate utility of all students, denoted by $V(s)$. We continue to assume that employers are able to observe the standard of the school and the type of an applicant i. Therefore, for any given standard s, individual choices of students are still determined by (2) and (3) and wages are still given by (4), where s_i is replaced by s.

Alternatively, one could assume that type i is not observable anymore when classes are merged. In such a modelling, type would be linked to the school an individual attends, rather than being a characteristic of the person. As explained above, we consider this interpretation to be too narrow. Moreover, coupling type to school would introduce an additional effect of a change in the schooling system. With a comprehensive school, employers would then have less information than with tracked schools. Merging schools would force wages to be equal across types, which would feed back into students' incentives and the school's standard setting policy. This effect would act on top of the impact of the need to set a uniform standard in the comprehensive school, rather than two standards tailored to tracked schools. Since we want to focus on the latter effect, and since it is rather plausible that the type is linked to the person, we assume that type i is observable in both schooling systems.

Depending on the standard, the objective function V can be of one of four different forms. First, all students choose a performance equal to their initial ability γ_i, either because they meet the standard without cost or because they choose not to graduate. We denote the value of the school's objective function in this case by $\tilde{V}_0$. Since in this case every student of type i earns a wage equal to γ_i and has no cost, it follows $\tilde{V}_0 = \gamma_l(1-d) + \gamma_h d$. This case will not occur when the school chooses an optimal standard.

Second, only for the l-students the standard is binding, while the h-students choose performance γ_h. Using the wage (4) and the cost function (1), aggregate utility in this case is $\tilde{V}_l(s) = \left[s - \frac{1}{2a_l}(s-\gamma_l)^2\right](1-d) + \gamma_h d$.

Third, both types of students choose to graduate and have to incur effort costs to do so. Then performance of students of both types just meets the standard s. Hence, the school's objective is $\tilde{V}(s) = \left[s - \frac{1}{2a_l}(s-\gamma_l)^2\right](1-d) + \left[s - \frac{1}{2a_h}(s-\gamma_h)^2\right] d$.

Fourth, l-students perform at their initial ability γ_l, while h-students meet the standard. This yields the objective function $\tilde{V}_h(s) = \gamma_l(1-d) + \left[s - \frac{1}{2a_h}(s-\gamma_h)^2\right] d$.

Which one of these four cases applies depends on how large the standard s is compared to initial abilities γ_i and the maximal standards $\gamma_i + 2a_i$ students of type $i = l, h$ will satisfy. In Appendix A.I in the supplementary material, we give these conditions explicitly.

Notice that $\tilde{V}_l(s), \tilde{V}_h(s)$, and $\tilde{V}(s)$ are strictly concave in s. Moreover, observe that $\tilde{V}_l$ and $\tilde{V}_h$ are affine transformations of the objective functions V_l and V_h of the separated classes: $\tilde{V}_l(s) = (1-d)V_l(s) + d\gamma_h$ and $\tilde{V}_h(s) = dV_h(s) + (1-d)\gamma_l$. Consequently, in any interval where $V(s) = \tilde{V}_l(s)$ or $V(s) = \tilde{V}_h(s)$, the optimal standard is s_l^* or s_h^*, respectively. In an interval where $V(s) = \tilde{V}(s)$, the optimal standard s^* solves:

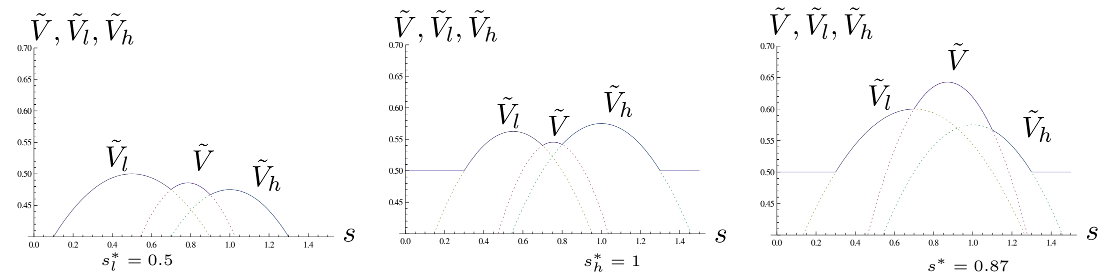
$$\frac{\partial \tilde{V}}{\partial s} = \left[1 - \frac{1}{a_l}(s-\gamma_l)\right](1-d) + \left[1 - \frac{1}{a_h}(s-\gamma_h)\right] d \overset{!}{=} 0 .$$

From this first order condition, we obtain:

$$\begin{aligned} s^* &= \frac{a_l a_h + d a_l \gamma_h + (1-d) a_h \gamma_l}{d a_l + (1-d) a_h} \\ &= \frac{d a_l}{d a_l + (1-d) a_h} s_h^* + \frac{(1-d) a_h}{d a_l + (1-d) a_h} s_l^* . \end{aligned} \tag{9}$$

In this case, the standard of a mixed class is a weighted average of the standards chosen in separated classes. The weights combine the population shares d and $(1-d)$ with the ability parameters a_h and a_l.

We will now analyze which of the three standards s_l^*, s_h^*, and s^* is the global optimum, that is, which of these three standards gives the highest welfare. As Figure 2 illustrates, each case is possible. The sub-figures 2(a), 2(b) and 2(c) are based on different parameter combinations, where the globally optimal standard is s_l^*, s_h^*, and s^* respectively.



(a) Parameter: $d = 0.5$, $\gamma_h = 0.7$, $\gamma_l = 0.1$, $a_h = 0.3$, $a_l = 0.4$

(b) Parameter: $d = 0.5$, $\gamma_h = 0.7$, $\gamma_l = 0.3$, $a_h = 0.3$, $a_l = 0.25$

(c) Parameter: $d = 0.5$, $\gamma_h = 0.7$, $\gamma_l = 0.3$, $a_h = 0.3$, $a_l = 0.4$

Figure 2: The objective function of the comprehensive school $V(s)$. $\tilde{V}_h$ ($\tilde{V}_l$) describes the value of the school's objective when h (l)-students exert effort to pass the exam and l (h)-students choose the effort level γ_l (γ_h). $\tilde{V}$ represents the school's objective where both types of students graduate with positive effort costs. $V(s)$ is the upper envelope of $\tilde{V}$, $\tilde{V}_l$ and $\tilde{V}_h$. For formal definitions, see (A.1) in the supplementary material.

The difference in welfare between standards s^* and s_l^* can be decomposed as

$$\begin{aligned} \tilde{V}(s^*) - \tilde{V}_l(s_l^*) &= \tilde{V}(s^*) - (1-d)V_l(s_l^*) - dV_h(s_h^*) + d\left[V_h(s_h^*) - \gamma_h\right] \\ &= \tfrac{da_h}{2} - \Delta. \end{aligned}$$

Here, we define $\Delta = (1-d)V_l(s_l^*) + dV_h(s_h^*) - \tilde{V}(s^*) > 0$ and use (6). The quantity Δ is the welfare loss caused by supplanting the common standard s^* for individually tailored standards for both types; we call it the compromise cost. The term $da_h/2 > 0$ measures the welfare loss caused by pushing the performance of h-type students from their optimal standard down to the initial endowment, as happens when the comprehensive school sets standard s_l^*. This cost is called the incentive cost for the h-type students. We obtain similarly

$$\begin{aligned} \tilde{V}(s^*) - \tilde{V}_h(s_h^*) &= \tfrac{(1-d)a_l}{2} - \Delta, \\ \tilde{V}_h(s_h^*) - \tilde{V}_l(a_l^*) &= \tfrac{da_h}{2} - \tfrac{(1-d)a_l}{2}, \end{aligned}$$

where $(1-d)a_l/2$ is the incentive cost for l-type students.

The compromise cost can be expressed as (see Appendix A.II in the supplementary material)

$$\Delta = \left[(a_h + \gamma_h) - (a_l + \gamma_l)\right]^2 \cdot \frac{d(1-d)}{2[da_l + (1-d)a_h]}. \tag{10}$$

One notices that this cost is the larger, the more total abilities $a_i + \gamma_i$, or equivalently type-specific optimal standards s_i^*, differ from each other. Moreover, the cost vanishes if one type represents the entire population ($d = 0$ or $d = 1$). Finally, the cost decreases if both a_i increase by the same amount. This reflects the fact that a high learning ability mitigates the cost of having to satisfy a standard which is not tailored to one's own need.

Pairwise comparison of the local maxima obtained by the three standards shows that $\tilde{V}(s^*)$, $\tilde{V}_l(s_l^*)$, or $\tilde{V}_h(s_h^*)$ is largest if Δ, $da_h/2$, or $(1-d)a_l/2$ is the smallest of the three cost. Since the cost terms are functions of the parameters of the model, this fully characterizes the global optimum provided that the optimal standard in each case indeed induces the student behavior which underlies the definitions of $\tilde{V}(s^*), \tilde{V}_l(s_l^*)$, and $\tilde{V}_h(s_h^*)$. This is the case, as the following proposition shows.

Proposition 1 *In the comprehensive school the chosen standard is:*

$$\begin{aligned}
s^* &= \tfrac{da_l}{da_l+(1-d)a_h} s_h^* + \tfrac{(1-d)a_h}{da_l+(1-d)a_h} s_l^* && \text{if } \Delta < \min\left\{\tfrac{da_h}{2}; \tfrac{(1-d)a_l}{2}\right\} \\
s_l^* &= a_l + \gamma_l && \text{if } \tfrac{da_h}{2} < \min\left\{\Delta; \tfrac{(1-d)a_l}{2}\right\} \\
s_h^* &= a_h + \gamma_h && \text{if } \tfrac{(1-d)a_l}{2} < \min\left\{\Delta; \tfrac{da_h}{2}\right\}.
\end{aligned} \tag{11}$$

Proof. See Appendix A.III in the supplementary material. □

From this proposition we can directly read the globally optimal standard. This is s^* if the parameters are such that the compromise cost is smaller than both incentive costs. If the incentive cost of one type of student is the smallest of the three cost, the comprehensive school chooses the standard which is optimal for the other type.

Figure 3 illustrates in which region of the parameter space each of the three local maxima is the global maximum. This figure is drawn in $a_h - a_l$-space, since the influence of these parameters on the optimal standard is most interesting to study. In this example the other parameters are fixed at $d = 0.5$, $\gamma_l = 0.3$ and $\gamma_h = 0.7$. In the graph we have inserted a dotted straight line, starting at $a_h = 0; a_l = 0.4$. Above this line inequality (8) holds. Below this line we have (7), such that labeling the l-type as low ability students is appropriate.

In the lower right region of the figure, labeled with s_h^*, the optimal standard in the comprehensive school is s_h^*, as in the tracked h-school. In this region the ability of

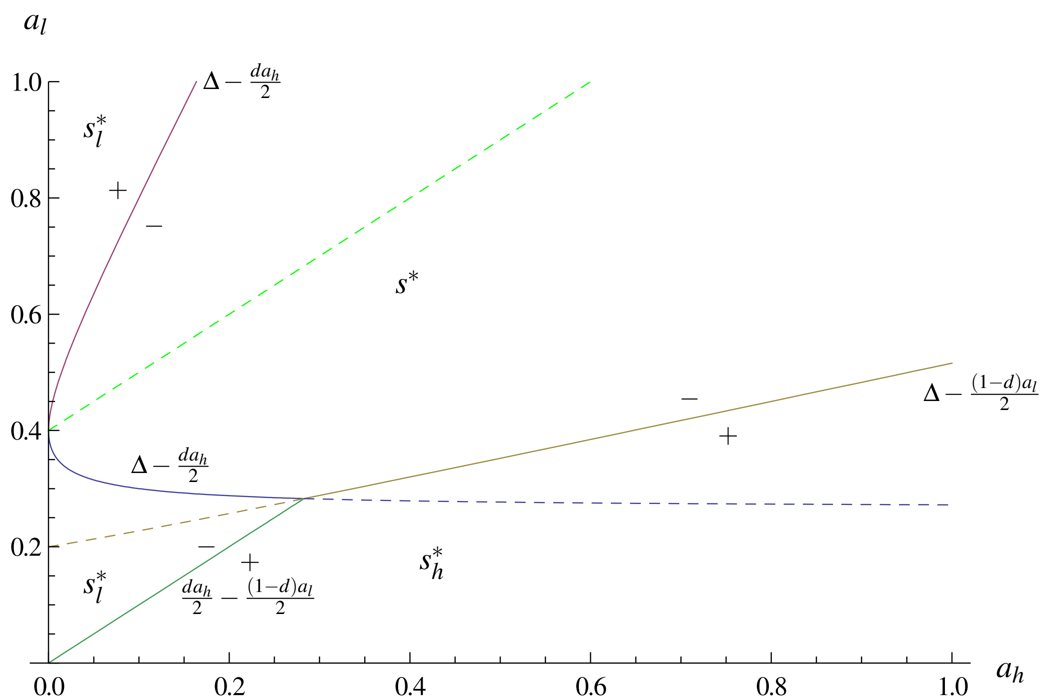


Figure 3: Parameter regions in $a_h - a_l$-space with different optimal standards s^*, s_l^*, and s_h^* in the comprehensive school, with $d = 0.5$, $\gamma_l = 0.3$ and $\gamma_h = 0.7$. Below (on, above) the dotted line starting at $(a_h = 0; a_l = 0.4)$ one has $a_l + \gamma_l < (=, >)\, a_h + \gamma_h$.

h-students is relatively high in both dimensions compared to the l-students. Therefore the teacher sets a standard tailored exactly to h-students, accepting that l-students will drop out since their incentive cost is low. In the central and upper right region, labeled s^*, the abilities of both types do not differ much such that the compromise cost is small, and hence the teacher sets the compromise standard s^*. Finally in the upper left and lower left regions, labeled with s_l^*, the school chooses the standard s_l^* for the l-students, whereas the h-students perform at γ_h. They do so for two different reasons. In the upper region with $a_l > 0.4$, where (8) holds, the standard is too high for the h-students and they drop out. In the lower part ($a_l < 0.4$, and (7) holds), the standard is so low that the h-students can meet it without effort cost. In both cases the incentive cost of h-students is low.

The next proposition provides comparative statics for the case where the optimal standard is s^*.

Proposition 2 *If the optimal standard in a comprehensive school is s^*, it increases in*

a_l, a_h, γ_l and γ_h. It increases (decreases) in d if $a_h + \gamma_h > (<)\, a_l + \gamma_l$.

Proof. See Appendix A.III in the supplementary material. □

As expected, the standard increases in both dimensions of ability of each type. Moreover, the standard increases in the share of the type of students i whose total ability, measured by $a_i + \gamma_i$, is larger.

IV Comparison of student performance

We now turn to the question whether two separated classes are preferable to the mixed class. It is important to distinguish between a comparison of utilities and a comparison of academic performances. Regarding utilities, no case is possible where students are better off in the mixed class than in separated classes, because in that case, s_l^* and s_h^* can be optimized separately. Hence, comparison of utilities is a straightforward application of the decentralization theorem by Oates (1972). Notice that the same observation would hold if we assumed that schools maximize academic performance or wages instead of students' utility. In such a model it would be immediate that in the comprehensive school performance of each type can only be worse than in tracked schools.

In contrast, as we will now show, in our model academic performance can also increase by mixing the classes. First, we consider for each type of students separately how their performance changes if classes are merged. For this comparison, observe that in the case where the optimal standard of the comprehensive school is s_l^* (s_h^*), the performance of the h (l)-students is γ_h (γ_l), and that the standard s^* is a weighted average of the standards chosen in the tracked schools. With this, Proposition 1 immediately leads to:

Proposition 3 *The performance of l (h)-students in the comprehensive school is*

(i) higher than in the tracked school if the comprehensive school chooses s^* and $a_l + \gamma_l < (>)\, a_h + \gamma_h$,

(ii) lower than in the tracked school if the comprehensive school chooses s^* and $a_l + \gamma_l > (<)\, a_h + \gamma_h$, or if the comprehensive school chooses s_h^* (s_l^*),

(iii) equal to the performance in the tracked school if the comprehensive school chooses s^* and $a_l + \gamma_l = a_h + \gamma_h$, or if the comprehensive school chooses s_l^* (s_h^*).

This proposition shows that our model can generate a positive peer group effect for low ability students, by which we mean the l-students, where (7) holds. In a comprehensive school, teachers will need to find a compromise between the standards tailored to individual student types. As long as low ability students are still willing to meet this standard, they will put in more effort than in the separated class. As a consequence one will observe higher test results on their part in the comprehensive school, even if there are no synergy effects from teaching diverse students together or behavioral norms inducing students to emulate classmates. Furthermore, the reduction of the standard for the h-students might be quite small when d and/or a_l are relatively large. Then, as (9) shows, the standard of the mixed class is close to the standard of the h-class. Given confounding influences, an empirical study might fail to find statistical significance of such a small impact.

The peer group effect obtains only for a subset of the parameter space. As is apparent from figure 3, the learning capacities of both types must not be too different. Otherwise, if one type finds it substantially easier to learn, the school will set optimal incentives for this type and put up with the fact that the other type stops graduating. Furthermore one can show that the $\Delta - \frac{(1-d)a_l}{2} = 0$ -curve shifts downwards if d decreases. Hence, a positive peer group effect for the low ability students is more likely when these students are more numerous. In this case the teacher of the comprehensive school puts more weight on their utility and therefore refrains from setting a standard which overburdens them.

Finally it may also happen that the comprehensive school sets the standard s_l^* tailored to the low ability students. This corresponds to the lower left region of Figure 3. In this case mixing classes leaves the performance of l-students unchanged and reduces the performance of h-students. However, this decline in performance might be very small, since learning capacity of h-students is in this case not very large anyway and since they continue to perform at their initial endowment γ_h. Therefore, although mixing classes obviously does not help in this case, the damage it inflicts is small.

We now turn to the effect of merging classes on aggregate performance. When the comprehensive school chooses a standard s_h^* or s_l^* tailored to one of the two groups, this group's performance is unchanged while the other group falls back on initial ability.

Hence, average performance will clearly decrease in this case. When the comprehensive school chooses the standard s^*, the outcome is less obvious, as the next proposition shows.

Proposition 4 *The average performance of students in a comprehensive school which chooses the standard s^* exceeds the average performance of students in tracked schools if and only if $a_l + \gamma_l < a_h + \gamma_h$ and $a_l > a_h$. That is:*

$$s^* > ds_h^* + (1-d)s_l^* \quad \Leftrightarrow \quad a_l + \gamma_l < a_h + \gamma_h \textit{ and } a_l > a_h \,.$$

Proof. See Appendix A.III in the supplementary material. □

This proposition shows that merging classes with heterogeneous students may increase overall academic performance, even when there are no spillover effects between types of students in learning or in preferences. This occurs when students with low initial ability have higher learning capacity than students with high initial ability. To understand that, consider how the standard is set in the comprehensive school. The teacher will trade off the net-loss incurred by l-students when the standard is increased above their optimal standard s_l^* against the net-loss incurred by h-students when the standard is decreased below s_h^*. Since the learning ability of l-students exceeds the learning ability of h-students, the net-loss of the latter increases faster than the net-loss of the former. Therefore the optimal standard, where marginal net-losses are equalized, is closer to s_h^* than to s_l^*. Hence the optimal standard in the mixed class is higher than the weighted average of the standards of the separated classes.

This kind of result is likely to be relevant in education systems where students are tracked early. In that case, it is particularly likely that the allocation to different tracks is mostly determined by the endowment of skills conferred by the family background. At the same time it is well possible that students with low endowment have not yet fully unfolded their potential and correspondingly find it easier to extend their knowledge. In the terminology of our model these students have high learning capacity a_l, but low endowment γ_l. If these students now attend a comprehensive school, average performance of students will increase. Both types of students find a relatively high standard acceptable, but they do so because of different reasons: One group starts with high initial ability and the others are eager to advance.

V Conclusion

In this paper we present a model comparing the choice of examination standards by tracked and untracked schools. The model distinguishes between initial ability and the capacity or willingness to extend ability. When setting the standard, the school or teacher takes the student's disutility of learning into account. Therefore, the resulting choices differ from the standards which maximize academic performance, which is the focus of PISA and similar studies.

Our findings show that in many cases a comprehensive school will enhance performance of low ability students or even enhance average performance compared to tracked schools with individual standards. In these cases performance of high ability students decreases, but this effect may be so small that it is insignificant in an empirical study. Our model therefore provides a foundation of peer group effects although we abstract from any synergy effect from teaching different student types together and from any kind of interdependent preferences.

Appendix

A.I The objective function of the comprehensive school

The four functional forms of the school's objective function mentioned in the text apply for different parameter constellations. This leads to the following definition of $V(s)$ with seven branches defined by parameter restrictions. Some of the restrictions are redundant, but are left for better understanding:

$$V(s) = \begin{cases} \tilde{V}_0 & [1] \quad \text{if } s \leq \gamma_{l,h} \\ \tilde{V}_0 & [2] \quad \text{if } s \leq \gamma_h;\ s > \gamma_l;\ s - \gamma_l > 2a_l \\ \tilde{V}_l(s) & [3] \quad \text{if } s \leq \gamma_h;\ s > \gamma_l;\ s - \gamma_l \leq 2a_l \\ \tilde{V}(s) & [4] \quad \text{if } s > \gamma_{l,h};\ s - \gamma_l \leq 2a_l;\ s - \gamma_h \leq 2a_h \\ \tilde{V}_h(s) & [5] \quad \text{if } s > \gamma_{l,h};\ s - \gamma_l > 2a_l;\ s - \gamma_h \leq 2a_h \\ \tilde{V}_l(s) & [6] \quad \text{if } s > \gamma_{l,h};\ s - \gamma_l \leq 2a_l;\ s - \gamma_h > 2a_h \\ \tilde{V}_0 & [7] \quad \text{if } s > \gamma_{l,h};\ s - \gamma_l > 2a_l;\ s - \gamma_h > 2a_h \end{cases} \tag{A.1}$$

In the first branch [1] the standard is too low to bind anybody, so all students just perform at the initial endowment. In the next branch [2] outcome is the same, but the standard is above the maximal standard the l-students are willing to satisfy and $u_{1l} < u_{0l}$, whereas h-students still graduate without effort cost. Branch [3] represents the situation where the l-students graduate by just meeting the standard whereas h-students still graduate with level γ_h. On branch [4], the standard is high enough to be binding also for h-students. Branches [5] and [6] differ depending on which group first refuses to satisfy the high standard and falls back to initial ability. In branch [5] this is true for the l-students and in branch [6] for the h-students. The last branch [7] shows a standard higher than anybody will accept to meet.

In Figure 2, $\tilde{V}_h$ ($\tilde{V}_l$, $\tilde{V}$) describes the value of the school's objective according to branch [5] ([3], [4]) of (A.1). Figure 3 illustrates in which branch of (A.1) the optimal standard is located. In the lower right region, labeled with $\tilde{V}_h$, the relevant branch is [5]. In the central and upper right region, labeled $\tilde{V}$, branch [4] of V contains the optimum. In the upper left region, with $a_l > 0.4$ and labeled with $\tilde{V}_l$, branch [6] of V is relevant. Finally, in the lower left region with $a_l < 0.4$, also labeled with $\tilde{V}_l$, branch [3] applies.

A.II Derivation of (10)

Using the definitions of the functions $\tilde{V}$, V_l, and V_h, we find

$$
\begin{aligned}
\Delta &= (1-d)V_l(s_l^*) + dV_h(s_h^*) - \tilde{V}(s^*) \\
&= (1-d)\left\{ s_l^* - \frac{1}{2a_l}(s_l^* - \gamma_l)^2 - \left[s^* - \frac{1}{2a_l}(s^* - \gamma_l)^2 \right] \right\} \\
&\qquad + d\left\{ s_h^* - \frac{1}{2a_h}(s_h^* - \gamma_h)^2 - \left[s^* - \frac{1}{2a_h}(s^* - \gamma_h)^2 \right] \right\} \\
&= (1-d)\left\{ s_l^* - s^* - \frac{1}{2a_l}\left[(s_l^* - \gamma_l)^2 - (s^* - \gamma_l)^2 \right] \right\} \\
&\qquad + d\left\{ s_h^* - s^* - \frac{1}{2a_h}\left[(s_h^* - \gamma_h)^2 - (s^* - \gamma_h)^2 \right] \right\} \\
&= (1-d)\left\{ s_l^* - s^* - \frac{1}{2a_l}\left[(s_l^* - \gamma_l) - (s^* - \gamma_l) \right]\left[(s_l^* - \gamma_l) + (s^* - \gamma_l) \right] \right\} \\
&\qquad + d\left\{ s_h^* - s^* - \frac{1}{2a_h}\left[(s_h^* - \gamma_h) - (s^* - \gamma_h) \right]\left[(s_h^* - \gamma_h) + (s^* - \gamma_h) \right] \right\}
\end{aligned}
$$

$$= \quad (1-d)(s_l^* - s^*)\left[2a_l - (s_l^* + s^* - 2\gamma_l)\right]\frac{1}{2a_l} + d(s_h^* - s^*)\left[2a_h - (s_h^* + s^* - 2\gamma_h)\right]\frac{1}{2a_h} .$$

Inserting the optimal standards from (5), we have furthermore

$$\Delta = (1-d)(s^* - s_l^*)^2 \frac{1}{2a_l} + d(s^* - s_h^*)^2 \frac{1}{2a_h} . \tag{A.2}$$

From (9), one has

$$\begin{aligned} s^* - s_l^* &= \frac{da_l}{da_l + (1-d)a_h}(s_h^* - s_l^*) , \\ s^* - s_h^* &= \frac{(1-d)a_h}{da_l + (1-d)a_h}(s_l^* - s_h^*) . \end{aligned}$$

Inserting these equations in (A.2), it follows

$$\begin{aligned} \Delta &= (s_h^* - s_l^*)^2 \cdot \left\{ \frac{1-d}{2a_l} \cdot \left[\frac{da_l}{da_l + (1-d)a_h}\right]^2 + \frac{d}{2a_h} \cdot \left[\frac{(1-d)a_h}{da_l + (1-d)a_h}\right]^2 \right\} \\ &= (s_h^* - s_l^*)^2 \cdot \frac{d(1-d)}{2[da_l + (1-d)a_h]} . \end{aligned}$$

Using the optimal standards from (5) once more, one arrives at (10).

A.III Proofs

Proof of Proposition 1

We need to consider all three possible optimalities and check if the needed constrains from (A.1) are satisfied. At first we consider

(i) s^* is optimal.
We need to show that if $\Delta < da_h/2$ and $\Delta < (1-d)a_l/2$ hold, s^* satisfies the conditions given in branch [4] of (A.1).

a) $s^* > \gamma_l$ is equivalent to $\frac{a_l[a_h + d(\gamma_h - \gamma_l)]}{da_l + (1-d)a_h} > 0$, which is satisfied in any case.

b) $s^* > \gamma_h$ is equivalent to $\frac{a_h[a_l - (1-d)(\gamma_h - \gamma_l)]}{da_l + (1-d)a_h} > 0$, which reduces to $a_l > (\gamma_h - \gamma_l)(1-d)$. Moreover, the inequality $\Delta < da_h/2$ is equivalent to $(1-d)\left[a_h - (a_l - \gamma_h + \gamma_l)\right]^2 < a_h[da_l + (1-d)a_h]$, or to $a_h\left[a_l(2-d) - 2(1-d)(\gamma_h - \gamma_l)\right] > (1-d)(a_l - \gamma_h + \gamma_l)^2$. This implies $a_l(2-d)-2(1-d)(\gamma_h-\gamma_l)>0$, which can be transformed into $a_l-(1-d)(\gamma_h-\gamma_l)>\frac{a_l d}{2}$. From this $a_l>(1-d)(\gamma_h-\gamma_l)$ and hence $s^*>\gamma_h$ follows.

c) $s^*-\gamma_l\leq 2a_l$ is, for $a_l\neq 0$, equivalent to $a_l\geq\frac{d(\gamma_h-\gamma_l)+a_h(2d-1)}{2d}\equiv A_l(a_h)$. The inequality $\Delta<(1-d)a_l/2$ is equivalent to $d[(a_h+\gamma_h-\gamma_l)-a_l]^2<a_l[da_l+(1-d)a_h]$, or to $a_l>\frac{d(a_h+\gamma_h-\gamma_l)^2}{a_h(1+d)+2d(\gamma_h-\gamma_l)}\equiv B_l(a_h)$. We show that $B_l(a_h)\geq A_l(a_h)$ so that $a_l>B_l(a_h)$ implies $a_l>A_l(a_h)$. To see this, observe that, with $(1+d)a_h+2d(\gamma_h-\gamma_l)>0$, we find that $B_l(a_h)\geq A_l(a_h)$ is equivalent to $(1-d)[a_h+d(\gamma_h-\gamma_l)]\geq 0$, which is true in any case.

d) $s^*-\gamma_h\leq 2a_h$ is, for $a_h\neq 0$, equivalent to $a_h\geq\frac{a_l(1-2d)-(1-d)(\gamma_h-\gamma_l)}{2(1-d)}\equiv A_h(a_l)$. The inequality $\Delta<da_h/2$ is equivalent to $(1-d)\left[a_h-(a_l-\gamma_h+\gamma_l)\right]^2<a_h[da_l+(1-d)a_h]$, or to $(1-d)(a_l-\gamma_h+\gamma_l)^2<a_h[(2-d)a_l-2(1-d)(\gamma_h-\gamma_l)]$. Since $\Delta<da_h/2$ implies $(2-d)a_l-2(1-d)(\gamma_h-\gamma_l)>0$ (see step **b**), $\Delta<da_h/2$ is equivalent to $a_h>\frac{(1-d)(a_l-\gamma_h+\gamma_l)^2}{a_l(2-d)-2(1-d)(\gamma_h-\gamma_l)}\equiv B_h(a_l)$. We show that $B_h(a_l)\geq A_h(a_l)$ so that $a_h>B_h(a_l)$ implies $a_h>A_h(a_l)$. The inequality $B_h(a_l)\geq A_h(a_l)$ is equivalent to $\frac{da_l[a_l-(1-d)(\gamma_h-\gamma_l)]}{(1-d)[a_l(2-d)-2(1-d)(\gamma_h-\gamma_l)]}\geq 0$. As shown in step **b**, we have $a_l-(1-d)(\gamma_h-\gamma_l)>0$ and $a_l(2-d)-2(1-d)(\gamma_h-\gamma_l)>0$. Hence $B_h(a_l)\geq A_h(a_l)$ for all $a_l>0$.

(ii) s_l^* is optimal.

Now we need to show that if $\Delta>da_h/2$ and $da_h/2<(1-d)a_l/2$ hold, s_l^* satisfies the conditions given in branch [3] or [6] of (A.1). We distinguish two cases, depending on whether (7) or (8) holds.

a) In the first case, where $a_h+\gamma_h>a_l+\gamma_l$, we show that branch [3] of (A.1) applies. The conditions $s_l^*>\gamma_l$ and $s_l^*-\gamma_l\leq 2a_l$ follow from $s_l^*=a_l+\gamma_l$. To see that also the condition $s_l^*\leq\gamma_h$ is satisfied, assume the contrary, i.e., $\gamma_h-a_l-\gamma_l<0$. Then, with (7), $a_h>a_h+\gamma_h-a_l-\gamma_l>0$, and hence $a_h^2>(a_h+\gamma_h-a_l-\gamma_l)^2$. Together with $\Delta>da_h/2$ this implies $a_h^2\cdot\frac{d(1-d)}{2[da_l+(1-d)a_h]}>\Delta>\frac{da_h}{2}$, hence $da_l<0$, a contradiction. Therefore, if (7) holds, $\Delta>da_h/2$ implies $s_l^*\leq\gamma_h$.

b) In the case $a_h+\gamma_h\leq a_l+\gamma_l$ the branch [6] of (A.1) applies. The conditions $s_l^*>\gamma_l$ and $s_l^*\leq 2a_l+\gamma_l$ follow directly from $s_l^*=a_l+\gamma_l$. The condition $s_l^*>\gamma_h$ is implied by $s_l^*-\gamma_h>2a_h$. We show that this latter condition follows from $\Delta>da_h/2$. Inserting $s_l^*=a_l+\gamma_l$, we can rewrite $s_l^*-\gamma_h>2a_h$ as $a_h<\frac{a_l+\gamma_l-\gamma_h}{2}\equiv C_h(a_l)$. Consider the denominator of $B_h(a_l)$ defined in step **(i.d)** above. From $a_h-a_l<\gamma_l-\gamma_h$ we have $a_l(2-d)-2(1-d)(\gamma_h-\gamma_l)>a_l(2-d)+2(1-d)(a_h-a_l)=a_ld+2(1-d)a_h>0$. Therefore $\Delta>a_h/2$ is equivalent to $a_h<B_h(a_l)$. We show that $B_h(a_l)<C_h(a_l)$, so

that $a_h < B_h(a_l)$ implies $a_h < C_h(a_l)$. The inequality $B_h(a_l) < C_h(a_l)$ is equivalent to $[a_l(-2+d)+2(1-d)(\gamma_h-\gamma_l)](a_l+\gamma_l-\gamma_h) < 0$. Since here $a_l+\gamma_l-\gamma_h > a_h > 0$, this is equivalent to $a_l(2-d)-2(1-d)(\gamma_h-\gamma_l) > 0$, which we just have shown to be true.

(iii) s_h^* **is optimal.**
$\Delta > (1-d)a_l/2$ is equivalent to $a_l < B_l(a_h)$, where $B_l(a_h)$ is defined in step **(i.c)** above. $s_h^*-\gamma_l > 2a_l$ is equivalent to $a_l < \frac{a_h+\gamma_h-\gamma_l}{2} \equiv D_l(a_h)$. We show that $B_l(a_h) < D_l(a_h)$. Knowing $a_h+a_hd+2d(\gamma_h-\gamma_l) > 0$ and $d < 1$, this is true. Hence $a_l < B_l(a_h)$ implies $a_l < D_l(a_h)$. The conditions $s_h^* > \gamma_h$ and $s_h^*-\gamma_h \leq 2a_h$ follow from $s_h^* = a_h+\gamma_h$. □

Proof of Proposition 2

Differentiating s^* from (9) we obtain:

$$\begin{aligned} \frac{\partial s^*}{\partial d} &= \frac{a_h a_l(a_h-a_l+\gamma_h-\gamma_l)}{(a_h(1-d)+a_l d)^2} && \text{(A.3)} \\ \frac{\partial s^*}{\partial \gamma_l} &= \frac{a_h(1-d)}{a_h(1-d)+a_l d} > 0 \\ \frac{\partial s^*}{\partial \gamma_h} &= \frac{a_l d}{a_h(1-d)+a_l d} > 0 \\ \frac{\partial s^*}{\partial a_l} &= \frac{a_h(1-d)\,[a_h+d(\gamma_h-\gamma_l)]}{[a_h(1-d)+a_l d]^2} > 0 \\ \frac{\partial s^*}{\partial a_h} &= \frac{a_l d\,[a_l-(\gamma_h-\gamma_l)(1-d)]}{[a_h(1-d)+a_l d]^2} && \text{(A.4)} \end{aligned}$$

(A.3) is positive (negative) if $a_h+\gamma_h > (<)\, a_l+\gamma_l$. (A.4) is positive if $a_l > (\gamma_h-\gamma_l)(1-d)$. As shown in the proof of Proposition 1, this is true if $s^* > \gamma_h$, which must be the case if s^* is the optimal choice. □

Proof of Proposition 4

From equation (9), we find that $s^* \gtreqless ds_h^*+(1-d)s_l^*$ is equivalent to:

$$(a_l-a_h)(s_h^*-s_l^*) \gtreqless 0\,. \qquad \text{(A.5)}$$

(i) 'if'. With $a_l+\gamma_l < a_h+\gamma_h$ we have $s_h^* > s_l^*$, and (A.5) is equivalent to $a_l \gtreqless a_h$. Hence, $a_l > a_h$ implies $s^* > ds_h^*+(1-d)s_l^*$.

(ii) 'only if'. From (A.5), with $a_l + \gamma_l < a_h + \gamma_h$, $a_l \leq a_h$ implies $s^* \leq ds_h^* + (1-d)s_l^*$. In the case $a_l + \gamma_l > a_h + \gamma_h$, we have $s_h^* < s_l^*$, and (A.5) is equivalent to $a_l \lesseqgtr a_h$. Since $\gamma_h > \gamma_l$, in this case $a_l > a_h$ must hold. Hence $s^* < ds_h^* + (1-d)s_l^*$. For $a_l + \gamma_l = a_h + \gamma_h$, we have $s_l^* = s_h^*$, and (A.5) implies $s^* = ds_h^* + (1-d)s_l^*$. $\square$

References

Argys, Laura M, Daniel I Rees, and Dominic J Brewer, "Detracking America's Schools: Equity at Zero Cost?," *Journal of Policy Analysis and Management*, 1996, *15* (4), 623–645.

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

Asirvatham, Jebaraj, Michael R. Thomsen, Rodolfo M. Nayga Jr, and Heather L. Rouse, "Do Peers Affect Childhood Obesity Outcomes? Peer-effect Analysis in Public Schools," *Canadian Journal of Economics/Revue canadienne d'économique*, 2018, *51* (1), 216–235.

Betts, Julian R, "The Impact of Educational Standards on the Level and Distribution of Earnings," *American Economic Review*, 1998, *88* (1), 266–75.

_ **and Jamie L Shkolnik**, "The Effects of Ability Grouping on Student Achievement and Resource Allocation in Secondary Schools," *Economics of Education Review*, 2000, *19* (1), 1–15.

Brunello, Giorgio and Daniele Checchi, "Does School Tracking Affect Equality of Opportunity? New International Evidence," *Economic Policy*, 2007, *22* (52), 782–861.

Chan, William, Hao Li, and Wing Suen, "A Signaling Theory of Grade Inflation," *International Economic Review*, 2007, *48* (3), 1065–1090.

Costrell, Robert M., "A Simple Model of Educational Standards," *The American Economic Review*, September 1994, *84* (4), 956–971.

—, "Can Centralized Educational Standards Raise Welfare?," *Journal of Public Economics*, 1997, *65*, 271–293.

Damiano, Ettore, Hao Li, and Wing Suen, "First in Village or Second in Rome?," *International Economic Review*, 2010, *51* (1), 263–288.

Duflo, Esther, Pascaline Dupas, and Michael Kremer, "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya," *American Economic Review*, 2011, *101* (5), 1739–74.

Epple, Dennis, Elizabeth Newlon, and Richard Romano, "Ability Tracking, School Competition, and the Distribution of Educational Benefits," *Journal of Public Economics*, 2002, *83* (1), 1–48.

Figlio, David N and Marianne E Page, "School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Inequality?," *Journal of Urban Economics*, 2002, *51* (3), 497–514.

Foureaux Koppensteiner, Martin, "Relative Age, Class Assignment, and Academic Performance: Evidence from Brazilian Primary Schools," *The Scandinavian Journal of Economics*, 2018, *120* (1), 269–325.

Friesen, Jane and Brian Krauth, "Sorting, Peers, and Achievement of Aboriginal Students in British Columbia," *Canadian Journal of Economics/Revue canadienne d'économique*, 2010, *43* (4), 1273–1301.

Hanushek, Eric A and Ludger Wößmann, "Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence Across Countries," *The Economic Journal*, 2006, *116* (510), C63–C76.

Hidalgo-Hidalgo, Marisa, "Tracking Can Be More Equitable than Mixing," *The Scandinavian Journal of Economics*, 2014, *116* (4), 964–981.

Jahanshahi, Babak, "Separating Gender Composition Effects from Peer Effects in Education," *Education Economics*, 2017, *25* (1), 112–126.

Kim, Taejong, Ju-Ho Lee, and Young Lee, "Mixing versus Sorting in Schooling: Evidence from the Equalization Policy in South Korea," *Economics of Education Review*, 2008, *27* (6), 697–711.

Liu, Xiaodong, Eleonora Patacchini, and Yves Zenou, "Endogenous Peer Effects: Local Aggregate or Local Average?," *Journal of Economic Behavior & Organization*, 2014, *103*, 39–59.

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

Oates, Wallace E, *Fiscal Federalism*, New York: Harcourt Brace Jovanovich, 1972.

Ostrovsky, Michael and Michael Schwarz, "Information Disclosure and Unraveling in Matching Markets," *American Economic Journal: Microeconomics*, 2010, *2* (2), 34–63.

Pekkarinen, Tuomas, "Gender Differences in Educational Attainment: Evidence On the Role of Tracking From a Finnish Quasi-experiment," *The Scandinavian Journal of Economics*, 2008, *110* (4), 807–825.

Popov, Sergey V and Dan Bernhardt, "University Competition, Grading Standards, and Grade Inflation," *Economic Inquiry*, 2013, *51* (3), 1764–1778.

Ruhose, Jens and Guido Schwerdt, "Does Early Educational Tracking Increase Migrant-Native Achievement Gaps? Differences-in-Differences Evidence Across Countries," *Economics of Education Review*, 2016, *52*, 134–154.

Thiemann, Kathrin, "Ability Tracking or Comprehensive Schooling? A Theory on Peer Effects in Competitive and Non-Competitive Cultures," *Journal of Economic Behavior & Organization*, 2017, *137*, 214–231.

World Health Organization, "Growing up Unequal: Gender and Socioeconomic Differences in Young People's Health and Well-being," Health Policy for Children and Adolescents, No. 7, 2016. WHO Regional Office for Europe, Copenhagen.

Wößmann, Ludger, "Institutional Determinants of School Efficiency and Equity: German States as a Microcosm for OECD Countries," *Jahrbücher für Nationalökonomie und Statistik*, 2010, *230* (2), 234–270.

Zubrickas, Robertas, "Optimal Grading," *International Economic Review*, 2015, *56* (3), 751–776.