The Optimal Timing of School Tracking

Giorgio Brunello*  Massimo Giannini†  Kenn Ariga‡

November 2003 VERY preliminary and incomplete

Abstract

1 Introduction

Most primary and secondary school systems in the developed world consist of an initial period of exposure to the same curriculum followed by diversification of curricula into separate tracks. Typically, diversification includes vocational and general or academic tracks, with allocation into tracks often based on previous performance and / or on ability tests (see Shavit and Muller, 1998, and Green, Wolf and Leney, 1999). This pattern is observed in continental Europe, where the timing of tracking can significantly differ. For instance, tracking starts relatively early after primary school in Germany and the Netherlands and later on, after junior secondary school, in France and Italy. Even when secondary schools do not formally diversify into general or vocational tracks, as in the US, grouping students in tracked classes for a subset of subjects is widespread, and the grouping is based on aptitude measures such as scores in standardized exams (see Epple, Newlon and Romano, 2002).

Differences in school design have recently been associated in the economic literature to differences in economic performance. Krueger and Kuman 2002,
for instance, have argued that the emphasis placed by Europe on specialized, vocational education may reduce the rate of technological adoption and lead to slower economic growth than in the United States, where the schooling system provides more general and comprehensive education. The broad idea is that general education is more suitable to induce (directed) technical change (see Acemoglu, 2000). Since general education is more flexible and versatile, it also encourages organizational change and the adoption of high performance holistic organizations in production (see Lindbeck and Snower, 2000, and Aghion, Caroli and Penalosa, 2000).

These contributions look at the effects of school design on technical and organizational change. It is natural to ask, however, whether and how these changes will affect in turn endogenous school design. Technical progress leads to skill depreciation, and the degree of obsolescence is likely to be higher the more specialized and tied to a specific set of techniques skills are. While skills learnt in vocational schools can be easily transformed into the corresponding occupations in the labor market, they are less flexible and transferrable than general skills (Shavit and Muller, 1998). As argued by Aghion, Caroli and Penalosa, 1999, organizational change is skill biased. Non hierarchical firms

"...rely on direct, horizontal communication among workers and on task diversification as opposed to specialization. They hence require multi-skilled agents, who can both perform varied tasks and learn from other agents'activities." (p.1651)

One clear effect of organizational change is the relative demand shift toward more general and versatile skills, which are better provided by general education.

School design clearly depends on a host of non-economic factors, including historical heritage. In this paper we try to answer this question by taking an economic perspective and by focusing exclusively on efficiency issues. By so doing, we are aware that we are only looking at one side of the problem. Nonetheless, we believe that our approach can be a useful complement to non-economic approaches as well as to other economic approaches which focus on the important distributional effects of school design.
We develop a simple model which determines the optimal timing of school tracking as the outcome of the trade off between the advantages of specialization, which would call for early tracking, and the costs of early selection, which would call instead for later tracking. In the model, the optimal tracking time is the time which maximizes total output net of schooling costs. We use this model to study how changes in the (exogenous) rate of technical progress and relative demand shifts toward more general skills affect the optimal tracking time as well as the allocation of students to schools.

Tracking is often associated to selection, and the key factor in the selection process is perceived ability\(^1\). In a world of imperfect information, selection conveys information to the labor market. Moreover, by grouping pupils of different abilities into classes, the overall effectiveness of schools in the production of human capital can be enhanced. As shown by Hoxby, 2001, peer effects have distributional effects but no efficiency implications if individual outcomes, such as human capital, are affected linearly by the mean of peers’ outcomes in that variable. Efficiency implications can only be drawn from models which are either nonlinear in peers’ mean achievement or in which other moments of the peer distribution matter (Hoxby, 2001, p.2)\(^2\).

In our model, the presence of nonlinear peer effects generates a positive "specialization" effect from tracking. In the absence of a countervailing factor, positive specialization would lead to immediate tracking. This factor is that the allocation of individuals to tracks is affected by noise in the test, and that the relative importance of noise is higher the earlier the test. Misallocation due to imperfect testing reduces both the quality of the signal offered by schools to the labor market and the peer effects in human capital formation. As remarked by Judson, 1998\(^3\).

---

1In Germany, "..the decision about school track is taken by both parents and the local educational authorities...but children’s measured ability remains the most important factor determining the selection process. This takes the form of a primary school recommendation for a secondary school track, generally based on a pupil’s marks in the core subjects of German and mathematics." (Schnepf, 2002)

2See for instance Epple and Romano 1998.

3See also Bedard, 1997 and Allen and Barnsley, 1993.
"..innate ability is measured with difficulty and with increasing clarity as education proceeds. Any test given will be a noisy signal, and the less education the person has had, the noisier the signal will be. Before primary school it is very difficult to discern levels of talent, but identification of talent is easier after a few years of primary school, still easier after high school, and so on." (p.340)

Therefore, the earlier selection is carried out, the higher the risk of misallocating individuals to the wrong track\(^4\). We call this the "noise" effect of tracking. The tradeoff between the positive "specialization" and negative "noise" effect generates an endogenous optimal time of stratification into tracks.

The importance of ability tracking for school performance has already been studied in the literature, most recently by Epple, Newlon and Romano, 2002. These authors, however, ignore noise in the selection process and treat both the threshold ability needed for allocation of pupils to tracks and the timing of stratification into tracks as exogenous parameters in order to study how varying these parameters modifies school performance.

Allocation of individuals to school types can be carried out either by prices (tuition fees) of by quantitative restrictions such as tests. Selection by test implies that individuals with a test score higher than the selected threshold are admitted to G schools and individuals with a lower score are allocated to V schools. Fernandez [1998] shows that allocation by tests should be preferred to allocation by prices when individuals are liquidity constrained.

In spite of the very simple structure of our model, its stochastic nature implies that we can provide relatively few analytical results. Therefore, we resort to calibration. By calibrating the key parameters of the model we can study how the optimal tracking time varies with the size of the peer effect and with the noise in the selection process. We show that there are regions in the space of parameters which yield internal solutions to the optimal problem. Therefore the observed allocation of schooling time to comprehensive and stratified schools can be efficient in the sense of maximizing total output net of schooling costs. We

\(^4\)While individuals can change tracks, this is not very frequent. Misallocation costs associated to peer effects remain even if we allow track switching.
also study how the efficient time of school tracking varies with key parameters, which include the exogenous rate of TFP growth and relative demand shift toward more general and flexible skills (NOT DONE YET, SORRY).

TO BE COMPLETED

2 The Setup

Consider a simplified economy with an exogenous number of individuals and job slots. Each individual lives for two periods. In the first (preliminary) period she goes to school and in the second period she is matched to a job slot supplied by a firm. The exogenous number of individuals is normalized to $S$. Due to minimum scale requirements as well as to limited resources, there is a given discrete number of schools $M$, run by the government, each with one teacher and $\frac{S}{M}$ students, where $S$ is the total number of students (see Lazear, 2001).

The assumption of public schools is quite accurate for most countries if we focus on education from primary to upper secondary level, but less accurate if we consider also tertiary education. While our model can be extended to include tertiary education, we prefer to focus our attention on primary and secondary education. In many countries this definition coincides with compulsory education, which justifies our assumption of an exogenous length of time at school. Notice that the percentage of individuals with at most an upper secondary education still covers an important share of the active population.

The cost $Z$ of running each school does not vary with school design. Depending on the design selected by the government, individuals spend a fraction of their schooling time $\tau \in (0, 1)$ in a comprehensive school, which provides the same curriculum to everybody, and the remaining fraction $(1 - \tau)$ in a stratified school, which tracks individuals into two different types of school, $G$ and $V$, each with its own specialized curriculum. In European countries, general and vocational schools start right after primary education (as in Germany and The

\footnote{In the US context, tracking consists of grouping pupils into different classes for some of the subjects taught (such as maths) within the same comprehensive school.}
Netherlands) and extend all the way to tertiary education, as documented by the OECD, which classifies tertiary education into types 5A and 5B, the former being general and theoretically based and the latter more practically oriented.

When \( \tau = 1 \) all \( M \) schools are comprehensive for the entire period of time. When \( \tau < 1 \) the \( M \) schools are comprehensive for time length \( \tau \) and are divided into \( MX \) vocational schools and \( M(1 - X) \) general schools for the rest of the time, where \( X \) is the percentage of pupils going to vocational schools. By assumption, there is no further stratification within each type of school.

Risk neutral individuals care only about (expected) wages and differ in their endowed ability, which cannot be observed by firms when recruitment takes place. While we can think of several types of ability, in this paper we focus only on cognitive ability, and assume that individuals differ in their endowment of this single type\(^6\). Firms only know the school the individual has graduated from. Since individual ability cannot be observed, each individual is paid her expected productivity. In this environment, each firm makes zero expected profits and the efficient outcome is produced by the school design which maximizes net output in the economy.

When individual utility depends only on expected wages after school and admission to G and V schools is free and left to individual choice, all individuals should enrol in schools G if the wage of graduates from these schools is expected to be higher than the wage gained by V graduates. We assume that allocation of students to schools is based on a noisy ability test: performance higher than or equal to a required standard qualifies the candidate for the general or higher-ability track and lower performance implies assignment to the vocational or lower-ability track. In practice, selection by test needs not be an entry exam, but can be based on the quality of the leaving certificate from the previous school, on orientation and evaluation by teachers and on selection during the first year after entry.

2.1 Schools

\(^{6}\)See Brunello and Giannini, 2004, for a discussion based on two ability types.
Using small letters for logarithms, let true ability $A \in (0, \infty)$ be log-normally distributed across individuals, and define $\alpha = \ln(A) \sim N(0, 1)$. Let observed log ability $\theta$ when the test takes place be related to true log ability by:

$$\theta = \alpha + \epsilon$$

(1)

where $\epsilon$ is an exogenous shock independent of $\alpha$ and normally distributed with mean zero and variance $b^2$. We capture the idea that the noise of the test increases the earlier the test is taken by letting

$$b = \mu (1 - \tau)$$

(2)

where $\mu$ is a suitable parameter.

Since $\alpha$ and $\epsilon$ are both normally distributed, the conditional density of $\alpha$ given $\theta$ is

$$\phi(\alpha|\theta) = \left( \frac{2\pi b^2}{1+b^2} \right)^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \frac{(\alpha - \theta)^2}{b^2} \right]$$

(3)

and the conditional mean is (see Anderson and Moore 1979)

$$E[\alpha|\theta] = \frac{\theta}{1+b^2}$$

(4)

Therefore, the expected true ability, conditional on unobserved ability, is higher the lower the variance of the noise. If the government sets the test standard $\theta^*$ to allocate individuals to schools, the expected true log ability of individuals allocated to school type is $E[\alpha|\theta \geq \theta^*]$ and $E[\alpha|\theta < \theta^*]$ in G and V schools respectively. Using the Law of Iterated Projections (Sargent 1979) we get

$$E[\alpha|\theta \geq \theta^*] = E[E[\alpha|\theta]|\theta \geq \theta^*] = \frac{1}{1+b^2} \int_{\theta^*}^{\theta^*} \theta \phi(\theta)d\theta = \frac{1}{1+b^2} E[\theta|\theta \geq \theta^*] = m_g$$

(5)

$$E[\alpha|\theta < \theta^*] = E[E[\alpha|\theta]|\theta < \theta^*] = \frac{1}{1+b^2} E[\theta|\theta < \theta^*] = m_v$$

(6)

Notice that $m_v$ is negative because the unconditional mean of $\alpha$ is equal to zero.

next we establish the following Remark.
Remark 1: The expected true log ability of pupils in G and V schools is increasing in the threshold $\theta^\ast$.

Proof. See Appendix.

An increase in the selection standard $\theta^\ast$ eliminates from G schools individuals in the lowest true ability group, who are allocated to V schools, where they belong to the highest ability group. Therefore, the expected true ability of either group increases (see Betts [1998]).

Each school combines individual ability with the effectiveness of teaching to produce human capital. Since by assumption the number and quality of schools and teachers is given, we posit that effectiveness vary with the average ability of the class (peer effect): the abler the class the more effective is instruction provided by a teacher of given quality. If an individual spends all her first period in a comprehensive school, her human capital at the end of the period is

$$H_c = A \exp \beta E(\alpha)$$

(7)

where $\exp \beta E(\alpha)$ is the peer effect, our measure of the effectiveness of teaching. The (log) human capital accumulated in this type of schools is

$$h_c = \beta E(\alpha) + \alpha = \alpha$$

(8)

Next consider schools stratified into tracks. Pupils in G schools have an observed ability $\theta^\ast$ higher than $\theta^\ast$. If they spend all their time in such schools their individual human capital is

$$h_g = \beta E[\alpha | \theta \geq \theta^\ast] + \alpha = \beta m_g + \alpha$$

(9)

where $m_g$ is the expected log time ability of students in school G. Similarly for V schools we have

$$h_v = \beta E[\alpha | \theta < \theta^\ast] + \alpha = \beta m_v + \alpha$$

(10)

\footnote{Zimmer and Toma, 2000; Hoxby, 2000; Zimmermann, 2003; Hanushek, Klain, Markman and Rivkin, 2001 is a non exhaustive list of recent contributions on peer effects.}
Students spend initial proportion $\tau \in (0, 1)$ of their time at school in a comprehensive school and the complementary proportion $(1 - \tau)$ in a stratified school. If schooling includes for example comprehensive and G schools, the individual *expected* log human capital at the end of the schooling period is

$$h_G = \tau h_c + (1 - \tau)h_g = \alpha + (1 - \tau)\beta m_g$$

(11)

and likewise for V schools, except from the fact that we allow vocational skills to depreciate at the rate $\delta g$, where $g$ is the rate of exogenous technical progress. The idea here is that vocational skills are less flexible and adjustable than general skills, so that they depreciate faster than general skills with the introduction of new technologies. As a convenient normalization, we set obsolescence of skills developed in general or comprehensive schools to zero. The human capital of an individual who has enrolled in a vocation school is

$$H_V = H_c^\tau (H_c (1 - \delta g))^{1-\tau}$$

so that, using logs and the approximation $\ln(1 - x) \simeq -x$, we obtain that the expected individual log human capital in vocational schools is

$$h_V = [\alpha + (1 - \tau)\beta m_c] - (1 - \tau)\delta g$$

(12)

where $\delta$ is a parameter. A higher rate of technical progress reduces the human capital accumulated in vocational schools, which is more closely linked to specific technologies.

In the second period graduates enter the labor market and are hired by firms, which only observe the school type (the same type if schools are fully comprehensive, general or vocational if schools are divided into tracks at some point in time). Firms have to infer ability by observing school type. Suppose that the graduate has spent all her education in a comprehensive school ($\tau = 1$). In this case her expected ability is

$$Eh_c = E(\alpha) = 0$$

(13)
If the graduate has spent part of her time in a comprehensive school and part in a G school, her expected ability is

\[ Eh_g = E(h_G|\theta \geq \theta^*) = (1 - \tau)\beta m_g + E(\alpha|\theta \geq \theta^*) = [1 + (1 - \tau)\beta] m_g \]  

(14)

because ability is time invariant and firms know that the graduate must have ability higher than \( \theta^* \) to qualify for general schools. Similarly, for graduates of V schools we have

\[ Eh_v = [1 + (1 - \tau)\beta] m_v - (1 - \tau)\delta g \]  

(15)

Casual observation of schooling around the world suggests that primary education and often lower secondary education are comprehensive, with tracking starting later on. In principle, however, we could have tracking from the start followed by a period of comprehensive schooling. Suppose for instance that tracking in G or V schools last for period \( (1 - \tau) \), followed by comprehensive schooling for the remaining period \( \tau \). Assuming that firms have information on the entire school curriculum, expected human capital would be as in (14) and (15), and so would be depreciation. The only key difference would be that noise and misallocation in selection is higher in the system were tracking starts earlier on.

### 2.2 Firms

The economy is populated by a given number \( Z \) of identical firms, which produce output by using two types of jobs or tasks, a "G" and a "V" task. General tasks are filled by individuals with general education and vocational tasks are filled by individuals with vocational education. For convenience we normalize \( Z = S = 1 \). The Cobb Douglas production technology is given by

\[ y = \lambda (n_g + Eh_g) + (1 - \lambda) (n_v + Eh_v) \]  

(16)

where \( y \) is log real output and \( n \) is log employment. Profit maximization yields

\[ w_g = \ln \lambda + y - n_g \; ; \; w_v = \ln(1 - \lambda) + y - n_v \]  

(17)
where \( w \) is the log wage rate. Relative wages in this economy satisfy the following condition

\[
w_g - w_v = \ln \frac{\lambda}{(1 - \lambda)} + n_v - n_g
\]  

(18)

Following Katz and Murphy [1992], \( \ln \frac{\lambda}{(1 - \lambda)} \) measures relative demand shifts in log quantity units. A demand shift toward more general tasks (a higher value of \( \lambda \)) can be met either by an increase in relative wages or by an increase in the relative supply of general skills or finally by a combination of both. Relative supply depends on the selection threshold, \( \theta^* \), and on the optimal timing \( \tau \), which are set by the government to maximize net output.

2.3 The Optimal Policy

When schools are comprehensive (\( \tau = 1 \)), graduates have the same expected human capital and can fill indifferently either task. Since perfect competition in the labor market guarantees that \( w_g - w_v = 0 \), relative employment is simply

\[
n_g - n_v = \ln \frac{\lambda}{(1 - \lambda)}
\]  

(19)

Labor supply is defined by

\[
\ln(N_g + N_v) = 0
\]

(20)

Therefore \( n_g = \ln \lambda \) and log output \( y \) is

\[
NY_c = y = \lambda \ln \lambda + (1 - \lambda) \ln(1 - \lambda) - \ln(MZ)
\]

(21)

With selection, there are \( 1 - \Phi(\theta^*) \) G graduates and \( \Phi(\theta^*) \) V graduates, and total net output can be re-written as

\[
NY_s(\tau) = \lambda \ln[1 - \Phi(\theta^*)] + (1 - \lambda) \ln \Phi(\theta^*) + [1 + (1 - \tau)\beta] [\lambda m_g + (1 - \lambda) m_v] \\
- (1 - \lambda) (1 - \tau) \delta g - \ln(MZ)
\]

(22)

The government maximizes net output by selecting the optimal values of \( \tau \) and \( \theta^* \). The first order conditions are
Remark 1 implies that an internal solution for the threshold $\theta^*$ exists if

$$-\frac{\lambda \phi}{1 - \Phi} + \frac{(1 - \lambda) \phi}{\Phi} < 0$$

This condition can be rewritten as

$$\frac{\lambda}{1 - \lambda} > \frac{1 - \Phi}{\Phi} = \frac{N_g}{N_v}$$

which implies from (18) that $w_g - w_v > 0$. Therefore, it is efficient to allocate some of the students to G schools and the rest to V schools if the more productive students in G schools are paid higher wages. We use this result to establish the following Lemma

**Lemma 1:** The moving average $[\lambda m_g + (1 - \lambda) m_v]$ is positive.

**Proof.** See Appendix. ■

This Lemma implies that, at the optimal value of the selection threshold, a linear combination of expected abilities from both tracks is higher than the expected ability from a comprehensive school, which is equal to zero by definition. We call this the "specialization effect" of tracking.

The first order condition with respect to $\tau$ is composed of four terms: the first term is negative and captures the fact that a later tracking reduces the gains from specialization. The second term is positive because later tracking is associated with lower depreciation of vocational skills; the last two terms capture among other things the changes in the conditional distribution of $\theta$ as

$$\chi_\tau(\tau, \theta^*, \lambda, g, \mu) = -\beta [\lambda m_g + (1 - \lambda) m_v] + (1 - \lambda) \delta g$$

$$+ [1 + (1 - \tau) \beta] \frac{\mu b}{1 + b^2} \left[ \lambda \frac{\partial m_g}{\partial \tau} + (1 - \lambda) \frac{\partial m_v}{\partial \tau} \right] + \left[ 1 - \frac{\lambda}{1 - \Phi} \left[ 1 - \frac{E[\theta^2 | \theta \leq \theta^*]}{1 + b^2} \right] \right] = 0$$

(23)

$$\chi_{\theta^*}(\tau, \theta^*, \lambda, g, \mu) = -\frac{\lambda \phi}{1 - \Phi} + \frac{(1 - \lambda) \phi}{\Phi} + [1 + (1 - \tau) \beta] \left[ \lambda \frac{\partial m_g}{\partial \theta^*} + (1 - \lambda) \frac{\partial m_v}{\partial \theta^*} \right] = 0$$

(24)
\( \tau \) varies and can take either sign. In the absence of noise, \( \mu = 0 \) and (23) boils down to

\[
- \beta \{[\lambda E[\theta | \theta \geq \theta^*] + (1 - \lambda)E[\theta | \theta < \theta^*]]\} + (1 - \lambda) \delta g = 0
\]  

(26)

Without skill depreciation the left hand side is negative and optimal \( \tau \) is equal to zero: in the absence of noise and depreciation, the positive effects of specialization prevails and tracking starts from the beginning of the schooling period. On the other hand, in the absence of peer effects \( (\beta = 0) \) the left hand side is positive and the optimal timing is \( \tau = 1 \).

Assuming that the second order conditions for the maximum hold, we can establish the following

**Proposition 1:** When an interior solution \( (\tau^*, \theta^*) \) exists, the effect of an acceleration in the rate of TFP growth \( g \) on the optimal time of stratification \( \tau^* \) is positive.

**Proof.** See Appendix. ■

An acceleration of growth increases the depreciation of skills provided by vocational schools. The optimal government response consists of delaying stratification. Unfortunately, because of the complexity of (23) this is the only analytical result that can be derived from the model. We provide a more detailed description of the results by using calibration and numerical solutions.

### 3 Calibration

The numerical solution of the model requires that we calibrate parameters \( \beta, \delta \) and \( \lambda \). We calibrate these parameters for Germany, France, Italy and the UK. Due to the very limited evidence on peer effects, we set the value of parameter \( \beta \) at 0.2, within the range (0.15,0.4) suggested by Hoxby on the basis of empirical studies for the US. With a Cobb Douglas production function, \( \lambda \) is the share on the total wage bill of the wages paid out to workers in G jobs. Therefore

\[
\lambda = \frac{W_g N_g}{WN}
\]
We associate G jobs to technicians and clerks and V jobs to service employees, skilled and semi-skilled blue collars. Total employment for European countries is taken from official Eurostat publications. We decompose $N$ into $N_g$ and $N_v$ by using the distribution of employment by occupations provided by the European Community Household Panel\(^8\). Gross annual wages are taken from ECHP 1994-2000. Using these sources, we obtain the values of $\lambda$ listed in the first column of Table 1.

Table 1. Calibrated values and baseline model. By country.

<table>
<thead>
<tr>
<th>Country</th>
<th>$\lambda$</th>
<th>$\tau$</th>
<th>$\theta^*$</th>
<th>$\mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td>0.526</td>
<td>0.52</td>
<td>0.98</td>
<td>0.64</td>
</tr>
<tr>
<td>France</td>
<td>0.553</td>
<td>0.72</td>
<td>0.89</td>
<td>0.84</td>
</tr>
<tr>
<td>UK</td>
<td>0.560</td>
<td>0.85</td>
<td>0.85</td>
<td>1.27</td>
</tr>
<tr>
<td>Italy</td>
<td>0.536</td>
<td>0.73</td>
<td>0.92</td>
<td>0.88</td>
</tr>
</tbody>
</table>

We calibrate $\delta$ by using a Mincerian wage equation and by assuming that the wage of blue collar and service employees is given by

$$W = \varsigma \exp (\beta X + \alpha H(1 - \delta g))$$

where $H$ is for schooling and $X$ is a vector of additional controls. We fit the log version of the above equation for service and blue collar employees using ECHP data for the 4 European countries after interacting years of schooling $H$ with the rate of TFP growth, $g$, which we take from Nickell and Layard [1999]. The estimated parameters suggest that $\delta$ is equal to 0.3.

We illustrate in Figures 1 to 4 how the optimal school design varies as $\beta$ and $\mu$ vary. In Figures 1 and 2 we plot the optimal values of $\tau$ and $\theta^*$ by using the values of the calibrated parameters for Germany and by allowing $\mu$ to vary between 0 and 3. In Figures 3 and 4 we set $\mu$ to 0.64, a value which would produce an internal solution for $\tau$, and allow $\beta$ to vary between 0 and 1. Figure 1 shows that, as $\mu$ increases from zero, the optimal value of $\tau$ also increases and becomes progressively close to its upper value, where schools are

---

\(^{8}\)The ECHP data - release 2003 contract 4/99 - are available at the Department of Economics of the University of Padova.
fully comprehensive. Figure 2 shows that the increase in $\tau$ is accompanied by a reduction in the optimal threshold $\theta^*$. Therefore, a noisier test reduces tracking and increase the percentage of slots available in general schools.

Figure 3 shows that tracking should begin at the start of the schooling system when the contribution of the peer effect to individual human capital is larger than 0.4. Interestingly, this is also the upper limit suggested by Hoxby, 2000, in her empirical review. As expected, a higher value of the peer effect also increases the selectivity of the school system (Figure 4).

The calibration of $\beta$, $\lambda$ and $\delta$ leaves two endogenous variables, $\tau$ and $\theta^*$, and an additional parameter, $\mu$, which measures the relative variance of the noise in the test with respect to the variance of true talent, $\alpha$. Clearly, it is very difficult to pin down this parameter. Rather than trying to do this, we compute the baseline solution of the model by assuming that the actual value of $\tau$ in our sample of countries is equal to the optimal value and by deriving $\mu$ as a result. To put it differently, rather than solving the model for $\tau$ as a function of $\mu$, we do the opposite, by imposing the condition that the optimal and actual values of $\tau$ coincide. For each country, the actual value of $\tau$ is computed as the ratio of the total years of schooling spent in a comprehensive system over the total years of schooling from primary school to upper secondary education. Based on the OECD definition of schooling levels in different countries, we obtain the numbers in columns 2 of Table 1.

Using these values and the other parameters, we solve the model for $\theta^*$ and $\mu$. The results are shown in the last two columns of Table 1. In Germany stratification starts earlier than in the UK. Since we are imposing that the size of the peer effect is the same across countries, the observed difference in the timing of stratification can only be accounted for by the lower variance of noise in German selection procedures.

TO BE COMPLETED
4 Appendix

- Proof of Remark 1:

\[
\frac{\partial m_g}{\partial \theta^*} = \frac{1}{1 + b^2} \phi(\theta^*) \left[ E [\theta \mid \theta \geq \theta^*] - \theta^* \right]
\]

is positive because the expression within brackets is positive. Similarly

\[
\frac{\partial m_v}{\partial \theta^*} = \frac{1}{1 + b^2} \phi(\theta^*) \left[ \theta^* - E [\theta \mid \theta \leq \theta^*] \right]
\]

is also positive.

- Proof of Lemma 1: the expression \( \lambda m_g + (1 - \lambda) m_v \) > 0 can be written as

\[
\lambda \int_{\theta^*}^{\theta^*} \int f(\theta) d\theta + (1 - \lambda) \int_{\theta^*}^{\theta^*} \int f(\theta) d\theta > 0
\]

Adding and subtracting from the left hand side of the above expression \( \lambda \int_{\theta^*}^{\theta^*} \int f(\theta) d\theta \) and using the facts that \( E(\theta) = 0 \) and \( m_v < 0 \), we can rewrite it as follows

\[
(1 - \lambda) [1 - \Phi(\theta^*)] < \lambda \Phi(\theta^*)
\]

which corresponds to (25) in the main text.

- Proof of Proposition 1: Ignoring for the sake of illustration parameters \( \delta \) and \( \mu \), total differentiation of the first order conditions yields

\[
\chi_{\tau \tau} \frac{\partial \tau}{\partial g} + \chi_{\tau \theta^*} \frac{\partial \theta^*}{\partial g} = -\chi_{\tau g} \frac{\partial g}{\partial \tau} - \chi_{\tau \lambda} \frac{\partial \lambda}{\partial \tau}
\]

(27)

\[
\chi_{\theta^* \tau} \frac{\partial \tau}{\partial g} + \chi_{\theta^* \theta^*} \frac{\partial \theta^*}{\partial g} = -\chi_{\theta^* g} \frac{\partial g}{\partial \theta^*} - \chi_{\theta^* \lambda} \frac{\partial \lambda}{\partial \theta^*}
\]

(28)

so that by Cramer’s rule we obtain

\[
\frac{\partial \tau}{\partial g} = \frac{-\chi_{\tau g} \chi_{\theta^* \theta^*} + \chi_{\theta^* g} \chi_{\tau \theta^*}}{\Delta}
\]

(29)
where

$$\Delta = \chi_{\tau\tau} \chi_{\theta^* \theta^*} - \chi_{\tau \theta^*} \chi_{\theta^* \tau}$$  \hspace{1cm} (30)$$

is positive if the second order conditions for a maximum hold. The second order conditions also imply that \( \chi_{\tau\tau} < 0 \) and \( \chi_{\theta^* \theta^*} < 0 \). Moreover \( \chi_{\theta^* g} = 0 \) and \( \chi_{\tau g} > 0 \), which guarantee the result.
References


[10] Hanushek, Klain, Markman and Rivkin

[12] Katz and Murphy, 1992,

[13] Lazear, E., 2001,


[15] Nickell and Layard, 1999


Figure 1: Changes in $\tau$ as $\mu$ increases

Figure 2: Changes in $\theta^*$ as $\mu$ increases
Figure 3: Changes in $\theta^*$ as $\beta$ increases
Figure 4: Changes in $\tau$ as $\beta$ increases