Efficiency and Equity of Education Tracking

A Quantitative Analysis*

Suzanne Bellue†

Lukas Mahler ‡

April 21, 2023

Click here for the latest version

Abstract

We study the long-run aggregate and distributional effects of school tracking – the allocation of students to different types of schools – by incorporating school track decisions into a general-equilibrium heterogeneous-agent overlapping-generations model. The key ingredient in the model is the child skill production technology, where a child's skill development depends on her classroom peers and the instruction pace in her school track. We show analytically that this technology can rationalize reduced-form evidence on the effects of school tracking on the distribution of child skills. We calibrate the model to data from Germany, a country with a very early and strict school tracking policy. Our model suggests that eliminating the parental influence on the school track choice that arises purely from own-track preferences improves social mobility while keeping aggregate output constant. An education reform that postpones the tracking age from age 10 to age 14 generates even larger improvements in intergenerational mobility. However, these come at the cost of efficiency losses in aggregate economic output. The size of these losses depends on the design of the instruction levels in each school track.

*We are indebted to Antonio Ciccone, Michèle Tertilt, and Minchul Yum for their continued and invaluable support and guidance in this project. In addition, we are very grateful to seminar and conference participants at the University of Mannheim for helpful discussions and suggestions. Suzanne Bellue gratefully acknowledges financial support from the German Academic Exchange Service (DAAD) and the German Research Foundation (through the CRC-TR-224 project A03). Lukas Mahler gratefully acknowledges the financial support from the German Science Foundation (through the CRC TR 224 Project A04) and the SFB 884 Political Economy of Reforms.

†University of Mannheim (suzanne.bellue@uni-mannheim.de)

‡University of Mannheim (lukas.mahler@gess.uni-mannheim.de)

1 Introduction

School tracking, which involves the allocation of students into physically distinct types of schools that differ in the curriculum taught, intensity, and length, is a common feature of education policy in OECD countries.[1] While school tracking is designed to improve teaching efficiency, it may hamper equality of opportunity in access to education. The rationale behind tracking is that grouping children according to ability creates more homogeneous peer groups and allows for tailored instruction levels and curricula, leading to improved educational outcomes (Bonesrønning et al., 2022; Duflo et al., 2011). However, because the ability of young children is imperfectly observed (Hanushek and Wössmann, 2006), introducing early tracking increases the risk of misallocation and the influence of parental background on the track decision. As a result, early tracking may impair mobility in educational and labor market outcomes across generations (Falk et al., 2021; Dustmann, 2004; Meghir and Palme, 2005; Pekkala Kerr et al., 2013). For that reason, the timing of school tracking is a recurrent issue in the public and academic debate about education reforms in countries with a strict and early tracking regime, such as Germany.[2]

We argue that the aggregate effects of changing the timing of school tracking depend on the efficiency-equity trade-off that results from improving teaching efficiency through homogeneous peer groups and letting parents choose which track to assign their children under imperfect information about their children's skills. However, providing a quantitative assessment of the long-run aggregate, distributional and inter-generational effects of the timing of school tracking policies requires a macroeconomic model of mobility with an understanding of how parents will optimally adjust their track assignation choices in response to changes in the uncertainty about their child's skills and in the composition of the peers. Macroeconomic models of mobility provide a useful environment to consider such effects but have so far largely ignored how the development of child skills is during schooling years.

To empirically underpin our modelization of child skills formation during schooling years and parental track choices, we first document the evolution of children's test scores and track assignation using the German National Educational Panel Study (NEPS) (Blossfeld et al., 2011). We find that students' test score ranks vary across years, suggesting that ability, measured by test scores, is imperfectly observed. At age 4, children's standardized test scores differ by parental socioeconomic group, and the difference persists until the end of

[1] An overview about school tracking policies in OECD countries is given in Chapter 2 in OECD (2013).

[2] There is substantial variation in the timing of school tracking across OECD countries. While in some countries, such as Germany and Austria, tracking occurs already at the age of 10, in other countries, like the US and UK, do not track at all during secondary school.

schooling. The school track choice happens at the age of 10 and is largely predicted by test scores and parental socioeconomic groups. Second, we see that most of the time, parents follow teachers' track choice recommendations. However, some of them deviate, most often in favor of their own track. Deviators are more likely than other students to then fail. We interpret this fact as parents having a form of preference for their own track.

As assessing the long-run effects of the timing of school tracking policies requires taking into account the effects of tracking on the educational outcomes of children, but also how these outcomes translate into labor market outcomes and outcomes across generations, we use a general equilibrium life-cycle Aiyagari framework of overlapping generations in which parents care about their offspring in the tradition of Becker and Tomes (1986).[3] Child skills during the schooling years evolve according to their parents' socioeconomic group, to the instruction pace in their track, and to their peers' average skills, and are subject to uninsurable skill shocks. The model is tailored to fit the German Education System, where children are tracked into two school tracks at the age of 10 based on a decision by the parents. As in the data, the track decision may be influenced by parental preferences for children to follow in their own education steps. While only one track directly facilitates access to college education, we allow for second-chance opportunities as children can decide to switch tracks after secondary school. Going to college incurs psychic costs, which are a function of child skills as well as time costs relative to non-college education. End-of-school child skills translate into adult human capital, which evolves stochastically over the working life and determines, together with the tertiary education decision, the labor earnings. The distribution of human capital across college and non-college workers affects prices, which in turn affects the school track choice. Finally, households can save into a non-state-contingent asset subject to life-cycle borrowing constraints. When children become independent, parents can also make a non-negative inter-vivos transfer.

The model builds around a parsimonious theory of how a child's skills are developed during school years. Going to a school that belongs to a particular school track affects children's skills directly through interactions with peers at their school and the pace of instruction that is taught in that school. Every child is assumed to have an ideal instruction pace at which she learns best. However, there can only be one instruction pace per school

[3]This is hard, if not impossible, to do in a purely reduced-form way, not only because of its demands on data, but also because a change in the allocation of children across tracks and, consequently, a change in the allocation of workers across skill levels, may entail general equilibrium effects. For example, suppose the share of children who are allocated to an academic track school increases substantially in the long run. In that case, the price of academically skilled labor in the economy should decrease. This, in turn, makes an academic track school less attractive, which affects the share again.

track, which is set endogenously by the policymaker. We show analytically that, under linear direct peer effects and complementarity assumptions between own skill and instruction pace, this gives rise to efficiency gains from tracking in terms of improving aggregate end-of-school skills. Indeed, absent any unforeseeable shocks to child skills, an optimal tracking policy should perfectly stratify children according to their skills as early as possible. In the presence of skill shocks, however, due to the risk of misallocation, it can be optimal to postpone tracking, even from an efficiency point of view. Finally, the theory implies that not all children gain from tracking and that the losses are often concentrated in the track with the lower average skill level. Thus, our child skill formation technology rationalizes some of the most robust empirical findings regarding school tracking in the literature.[4]

The model is solved numerically, and the parameters are calibrated in two steps. First, we estimate the child skill formation technology parameters directly from German data on school children using a latent variable framework as in Cunha et al. (2010). In particular, we use the information on achievement test scores to measure child skills at different stages of their school careers. We use an instrumental variable strategy similar to Agostinelli et al. (2019) to account for measurement error. We then calibrate the remaining parameters to match a set of critical moments from representative German survey data. The model matches the data well, both in terms of aggregate moments and with respect to the distribution of child skills across school tracks and parental backgrounds, as well as the transitions through the education system. To test the model's validity, we investigate the effects of the initial school track on later-in-life economic outcomes for a set of children who are, in equilibrium, just at the margin between the two school tracks. Dustmann et al. (2017) argue that for such marginal children in Germany, the initial track choice is inconsequential for labor earnings later in life. Simulated data from our model suggest that children who go to different school tracks solely based on small differences in skills at the time of the track decision experience very similar lifetime economic outcomes, where children in an academic track school track earn around 2% higher lifetime labor income compared to similar children that did not go an academic-track school.

Notwithstanding this, our quantitative results show that the school tracking policy plays an important role in lifetime inequality across the population. In particular, variation in the initial school track alone can account for 15% of the variation in lifetime earnings and 19% of the variation in lifetime wealth. As in the data, parental education is, after child skills,

[4]For the case of Germany, see for instance Matthewes (2021) who shows that earlier tracking raises inequality in educational outcomes and Piopiunik (2014), who shows that low-achievers may be negatively affected by school tracking.

the second most crucial determinant of initial school track choice. We use our model to show that most of this effect comes from direct parental preferences for children to follow their education track rather than college tastes or knowledge about the deterministic influence of parental education on child skill development. The parental bias in school track choice gives rise to inefficiencies in the allocation of children across tracks. For example, a college-educated parent may push her child into an academic-track school even though her child's skills optimally suggest a vocational-track school. This harms her child's learning outcomes and affects average learning in that track as the instruction pace endogenously adjusts to the composition of skills in that track. We perform counterfactual experiments using our model that eliminate the parental bias in school track choice, for example, by introducing a strict skill threshold that governs school track allocation. Such a policy improves social mobility across generations with minor effects on cross-sectional inequality. The intergenerational income elasticity decreases by around 2-3%. However, while this policy improves the average learning outcomes of children, we highlight that aggregate GDP remains essentially unchanged. This is because the improvements in child skills are quantitatively minor (0.2%) in the first place and fade out over the remaining schooling career.

Finally, we use our model to study the long-run effects of an education reform that universally postpones the school tracking age by four years. Such a reform is often suggested in countries with traditionally early tracking systems, such as Germany, to improve equality of opportunity in access to academic education (Woessmann, 2013). We show that postponing the tracking age indeed improves social mobility as it leads to an 8-10% decrease in the intergenerational elasticity of income. However, this comes at the cost of around 3% of the GDP. The reason for this significant drop is that postponing tracking incentivizes most parents to send their children to academic track schools, regardless of their education. This results in substantial learning losses as secondary schooling becomes close to a fully comprehensive system. The learning losses incurred by foregoing four years of early tracking cannot be recuperated by efficiency gains coming from the fact that the late tracking decision is based on more complete information about children's skills. In addition, despite the larger share of graduates from academic track schools, the general equilibrium adjustment of prices for college human capital curbs the college wage premium, to the effect that the share of college-educated workers in the counterfactual economy increases by only 4%. A policy that exogenously fixes the instruction paces in each school track can go some way in attenuating the output losses from a postponed tracking policy while maintaining most of the social mobility gains.

Related Literature

This paper links several strands of the literature: the quantitative family-macroeconomics literature, the children's skill formation literature, and the school tracking literature.

First, we contribute to the quantitative family macroeconomics literature that studies determinants of the intergenerational transmission of economic status (Abbott et al., 2019; Caucutt and Lochner, 2020; Daruich, 2022; Jang and Yum, 2022; Fuchs-Schündeln et al., 2022; Fujimoto et al., 2023; Lee and Seshadri, 2019; Yum, 2022). Some of these studies incorporate a part of the educational system into their analysis, such as Abbott et al. (2019); Caucutt and Lochner (2020); Fuchs-Schündeln et al. (2022) who model high-school graduation choice. However, all of these studies except Fujimoto et al. (2023) focus on the United States, often concentrating on access to higher education and neglecting the importance of designing the (secondary) school system for macroeconomic outcomes. We explicitly focus on the secondary schooling system. Our paper is perhaps most closely related to to Fujimoto et al. (2023) who study the importance of free secondary schooling for misallocation driven by borrowing constraints in Ghana. Our contribution is to analyze a widespread education policy at the secondary school stage in developed countries: school tracking. In particular, we investigate the consequences of the school track choice and the age at which school tracking occurs for inequality and efficiency in a dynamic macroeconomic model. We thereby complement related research that focuses on the early, pre-school phases in a child's skill development (Daruich, 2022; Yum, 2022) and research that focuses on higher, post-secondary education (Abbott et al., 2019; Capelle, 2022).

Second, this paper builds on the literature on children's skill formation, which has described how children's skills evolve as a function of endowments, parental and environmental inputs, and recently also schooling inputs (see, for instance, Cunha and Heckman (2007); Cunha et al. (2010); Agostinelli et al. (2023, 2019)). Our main innovation relative to this literature is considering two forms of peer effects, which allows for rationalizing the empirical findings regarding school tracking. First, similar to Agostinelli (2018), we incorporate direct peer effects, which capture the idea that children benefit from high-quality peer groups. Second, following Duflo et al. (2011) 's evidence in Kenyan primary schools, we consider how the instruction levels adjust endogenously to the peer composition in schools of a particular track. More specifically, we assume that a child's optimal pace of instruction is unique and increases with her current skill level. Then, learning decreases monotonically with the distance between a child's optimal instruction pace and the one she is currently taught at. This parsimonious micro-funded model captures the main arguments about school tracking and allows us to evaluate the effects of delaying the tracking decision.

Third, this paper contributes to and builds on the literature that estimates the impact of early school tracking on efficiency and equity measures. An extensive empirical literature investigates the effects of age at school tracking on students' test scores and later outcomes. It either exploits temporal within-country variation in tracking practices (Meghir and Palme (2005), for Sweden; Aakvik et al. (2010), for Norway; Malamud and Pop-Eleches (2011), for Romania; Pekkala Kerr et al. (2013), for Finland; and Matthewes (2021); Piopiunik (2014) for Germany) or between-country variation with a difference-in-differences strategy (Hanushek and Wössmann, 2006; Ruhose and Schwerdt, 2016). Most studies suggest that earlier tracking raises inequality in educational outcomes and increases the effect of parental education on student achievement. Guyon et al. (2012) investigate an educational reform in Northern Ireland that led to a large increase in the share of students admitted to the elite track at age eleven. They find a strong positive overall effect of this de-tracking reform on the number of students passing national examinations at later stages and a negative effect on student performance in non-elite schools that lost their most able students. A notable exception is Dustmann et al. (2017), who use an individual-level instrumental variables strategy (the date of birth) and find no effect of track choice on educational attainment or earnings for German students at the margin between two tracks. While their result suggests that the misallocation of hard-to-assign students has little impact on their future outcomes, it does not rule out a potential adverse effect of early school tracking on the outcomes of non-marginal sub-groups of students, such as those from low-socioeconomic backgrounds.

The remainder of the paper is organized as follows. Section 2 provides empirical facts to guide our school-tracking modeling choices. Section 3 presents our model that uses a life-cycle Aiyagari GE framework of overlapping generations, and Section 4 helps build intuitions about the model mechanisms underlying school tracking using a parsimonious theory for the child skill formation. Section 5 explains how we estimate and parameterize the model. It also offers some validation exercises. In Section 6, we use the calibrated model to perform a series of counterfactual experiments to quantify the effects of different school tracking policy regimes. Finally, Section 7 concludes.

2 Data

In this section, we document the evolution of child skills along the parental socioeconomic background and school track dimensions in Germany that motivate our focus on school tracking, and that will also serve as inputs for the calibration of the quantitative model. We use the German National Educational Panel Study (NEPS), which comprises detailed longitudinal data on the educational process, acquired competencies, as well as the learning environment and context persons of six cohorts of children in nationally representative samples in Germany (Blossfeld et al., 2011). A key component of the information collected is regular standardized assessment tests of the children's competencies in areas such as mathematics, reading, sciences, vocabulary, or grammar, combined with specific wave weights. In addition, there is information about school track recommendations and the school track choice. Primary school teachers give these recommendations before transitioning to secondary school. They are based on reflecting on the child's achievement during primary school and the teacher's assessments.[5]

Table 1 shows that parents deviate from teacher recommendations toward their own education track. Research on school tracking has found that parents with higher socioeconomic status (SES) are more likely to send their child to an academic track school than parents with a lower socioeconomic status, even conditional on school performance or achievement test scores before the track decision (Falk et al., 2021). Consistently, we find that 74% of children from college-graduated parents receive a teacher recommendation for the academic track versus 43% of children from non-college-graduated parents.[6] In addition, Table 1 shows that around 23% of parents who themselves have a college education overrule a vocational recommendation. At the same time, 16% of non-college graduated parents overrule an academic recommendation. There may be multiple reasons behind these deviations. For example, parents may have more information about their child's skills than teachers. However, deviations are not symmetric, and we tend to see a bias toward parents' own education.

Parents may have several motivations that lead them to overrule teacher decisions more often when they differ from their own educational path. For instance, they may be able to better support their child in a track they are more familiar with. Nevertheless, among deviators, while children who are advised on the academic track do very well, children who are advised on the vocational track do poorly. 56% of children of non-college parents who deviated from the academic track belong to the top quartile of skills in G9. In contrast, only 20% of children of college parents who deviated from the vocational track belong to the top quartile of skills in G9. Thus, we would argue that parents' deviation when choosing the school track is partly due to a bias toward their education that is not motivated by their ability to support the child or their intrinsic knowledge of their skills.

[5] See also Appendix Section A.4 for more details on the tests.

[6] We define children from college parents if they have at least one of the parents with a college education.

Table 1: School Track Choice

	Recommendation (G5)	Shares	% deviate	% top 25% in G9 if followed	% top 25% in G9 if deviated
College Parents					
	Academic	74% (973)	6% (58)	34% (235)	51% (17)
	Vocational	26% (350)	23% (61)	31% (39)	20% (7)
Non-College Parents					
	Academic	43% (753)	16% (121)	24% (113)	56% (35)
	Vocational	57% (999)	12% (89)	25% (106)	23% (11)

Notes: This table provides information on school track choice by parental education and teacher recommendation. All observations are weighted. Source: NEPS, Cohort 3.

Second, we see that measures of a child's skills vary across academic years.[7] Table 2 shows the correlations of a child's math test score percentile rank across periods.[8] They are below one, suggesting that a child's skills are not perfectly observed at any point in time and are subject to variation. In addition, we see that early correlations are weaker than later ones. In particular, the correlation between Grades 1 and 4 is 0.59 and rises to 0.72 between Grades 5 and 9. We interpret these data facts as abilities being imperfectly observed and subject to shocks that may be larger in the earlier periods of a child's life.

Table 2: Rank-Rank Correlations

		Rank-Rank Correlation	Obs.
Cohort 2	G1-G4	0.59	3,116
	G1-G7	0.63	2,321
Cohort 3	G5-G9	0.72	5,311

Notes: This table provides the rank-rank correlation in math grades. Source: NEPS

Finally, in line with the literature, Table 3 shows that parental background influences child skills very early on. As of Grade K1, average child skills by parental education differ by

[7]We use mathematics grades as the main measurement of a child's raw skills.

[8]The sample comprises only children who took the tests in both Gardes. For this reason, we prefer unweighted estimates.

0.47 SD. This result is consistent with the literature that documents early skill level differences, which are a combination of parental investment and genetic components. Moreover, we see that skill differences by parental background tend to grow or stay stable over time, which is consistent with previous empirical studies.[9]

Table 3: Differences in Average Skills in Standard Deviation

	Grade	Difference by Parent's Education	Difference by School Track
Cohort 1	K1	0.47	
Cohort 2	G1	0.46	0.77
	G2	0.45	0.80
	G4	0.58	0.86
	G7	0.49	0.97
Cohort 3	G5	0.58	1.00
	G7	0.67	1.07
	G9	0.70	1.16
Cohort 4	G9	0.62	1.06

Notes: This table provides information on average differences in math grades in one standard deviation unit by parental background and school track over time. All observations are weighted. Source: NEPS

3 The Model

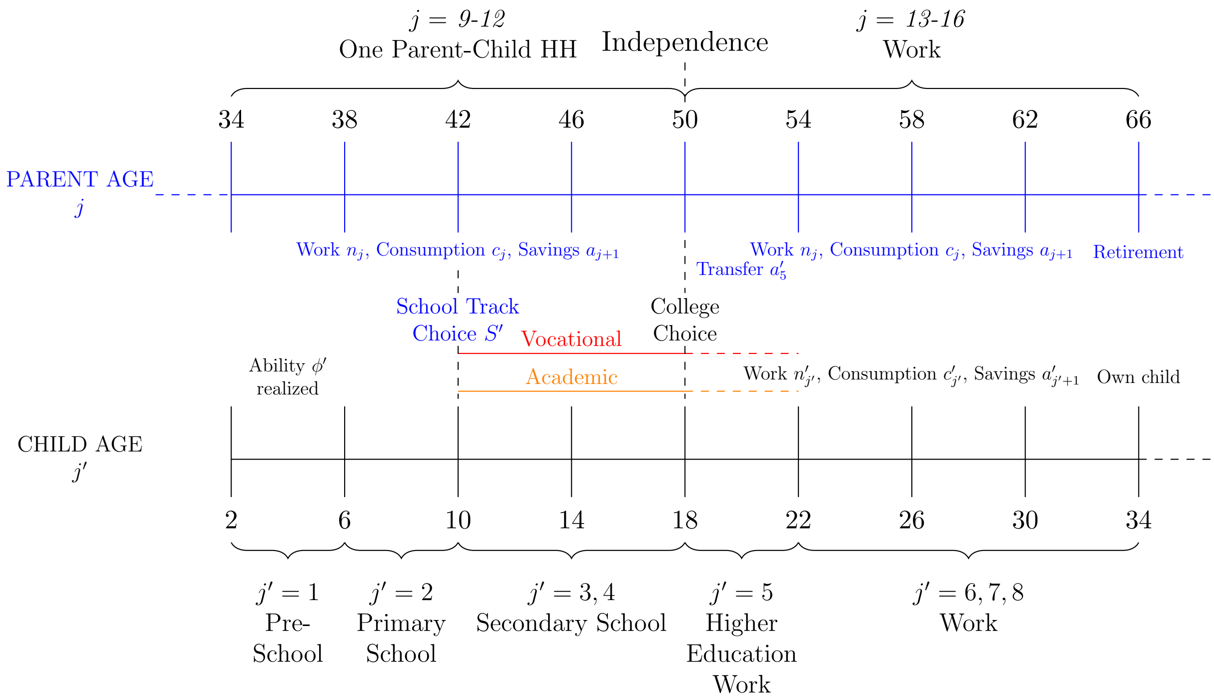
The model is designed to fit the German Education System and is guided by the facts documented above.[10] First, empirical variation in children's skills over time and parents' deviation from teacher recommendations motivate us to include age-specific skill shocks to the child skill formation and a parental preference for their education when deciding on the school track. Second, the parent's decision about their child's school track requires an overlapping-generations life-cycle model. We assume that time is discrete and infinite and that one model period, j, corresponds to the 4 years in between ages $[4j-2, 4j+2]$ in real life.[11] This frequency allows us to investigate meaningful variations in school tracking ages.

[9]See for instance Passaretta et al. (2022); Nennstiel (2022); Schneider and Linberg (2022) who investigate the NEPS data and find stable or growing socioeconomic status gaps in children's skills.

[10]An overview of the German Education System is given in Appendix A.3.

[11]We choose this perhaps unorthodox timing, such that children are 10 years old when parents make the secondary school track decision, which resembles reality in Germany.

Figure 1: Timeline of Life-cycle Events



The dynastic structure implies that there are 20 generations alive at every point in time. As in Lee and Seshadri (2019), we assume that there is a unit mass of individuals in each period. A life cycle can be structured into several stages, represented by j, as illustrated in Figure 1. For the remainder of the text, we will denote all child variables with primes. Since a generation is 32 years, the child of a parent who is in period j is in period $j' = j - 8$.

Each individual goes through 20 stages in life and starts with childhood, $j = 1, 2, 3, 4$, during which she makes no decisions, is educated by her parent, and goes to the school chosen by her parent. School tracking between academic and vocational schools happens in period $j = 3$. At age 18, in period $j = 5$, the child finishes school and becomes an independent adult. She decides how much to work and whether to pursue a college education or obtain a vocational/professional degree. Obtaining a college education substantially reduces the time available to work as it generally takes longer than obtaining a vocational degree. Over the following periods ($j = 6, 7, 8$), the individual works while not yet having children. From periods $j = 9$ to $j = 16$, the adult goes through the parenting years while also making consumption, savings, and labor supply decisions. During these periods, her human capital grows stochastically. Moreover, when her own child turns 18 and becomes independent, the parent decides on an inter-vivos transfer. In the remaining model periods $j = 17, ..., 20$, the individual is retired and makes consumption and savings decisions while earning retirement benefits. Everyone dies with certainty after model period $j = 20$, that is, at age 82.

3.1 Child Skill Formation

In period $j = 1$, a 2-year-old child enters into a one-parent household, equipped with an initial learning ability ϕ', which is imperfectly transmitted from her parent as in Yum (2022).[12]

$$\log \phi' = \rho_\phi \ \log \phi \ + \epsilon'_\phi, \qquad \epsilon'_\phi \sim \mathcal{N}(0, \sigma^2_\phi), \tag{1}$$

where ϵ_ϕ is an intergenerational shock. The learning ability translates into an initial child skill level

$$\theta'_1 = \log \phi'. \tag{2}$$

During the period of childhood, which consists of periods $j = 1,\ 2,\ 3,\ 4$, a child's skills are determined by past skills, household characteristics, and school inputs, as represented by the following equation:

$$\begin{aligned} \theta'_{j+1} &= g_j(\theta'_j,\ E,\ \bar{\theta}'_{j,S'}) \\ &= \kappa_{0,j} + \kappa_{1,j}\theta'_j + \kappa_{2,j}{\theta'_j}^2 + \kappa_{3,j}\bar{\theta}'_{j,S'} + \kappa_{4,j}\left(\theta'_j - \bar{\theta}'_{j,S'}\right)^2 + \kappa_{5,j}E + \eta'_{j+1}, \end{aligned} \tag{3}$$

where $\kappa_{2,j}$, $\kappa_{3,j}$, and $\kappa_{4,j}$ are set to zero in the pre-school period ($j = 1$), as they correspond to the school inputs. We allow for the explicit dependence of child skill development on parental education (E) measured by $\kappa_{5,j}$, which enables us to capture the effects of the household environment, as well as parental inputs, such as monetary investments in the child's skill development. The school input consists of two elements: linear and direct peer effects ($\kappa_{3,j}$) and the distance to the pace of instruction, measured by $\kappa_{2,j}$ and $\kappa_{4,j}$. The linear and direct peer effects capture the idea that children benefit from being surrounded by more able peers. The pace of instruction represents the tailored instruction levels and curricula and is written as a function of the average skill composition in a given track ($\bar{\theta}'_{j,S'}$).[13] The distance to the pace of instruction reflects the intuition that the closer a child is to the pace

[12]The learning ability captures genetic components and investments made by parents into their child's skill development during early childhood, infancy, and even in-utero. The importance of these early life stages as well as policy interventions targeted at investments during these years, has been the focus of the child skill development literature (see e.g. Heckman and Mosso (2014) for a review).

[13]In Section 4, we show that the pace of instruction that is set optimally by policymakers is a function of the average skills in the class. Here, we replace the dependence of child skills on the instruction pace with the average peer skills in school track S' and its squared term.

of instruction, the faster they learn.[14] In primary school, all children are in comprehensive schools ($j = 2$), and the average skills of peers $\bar{\theta}'_2$ correspond to the average skills in the cohort, assuming a representative classroom. At the beginning of secondary school ($j = 3$), children are assigned to either the academic or vocational track by their parents ($S' = A, V$). In periods $j = 3, 4$, the school component will depend on the composition of students in each of the tracks.

End-of-school skills , re-scaled with a constant factor ξ, determine the first adult human capital level, h_5:

$$h_5 = \xi \exp(\theta_5). \tag{4}$$

When entering adulthood ($j = 5$), individuals thus differ in their human capital h_5, their school track S, and their learning ability ϕ. In addition, we assume they receive a financial transfer from their parents, and thus differ in initial savings a_5.

3.2 Preferences, Labor Income and Social Transfers

We assume that the preferences of adult individuals over consumption and leisure take the following form:

$$u(c_j, n_j) = \frac{(c_j/q_j)^{1-\sigma}}{1-\sigma} - b\frac{n_j^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}}, \tag{5}$$

where q_j is an adult consumption-equivalent scale that depends on the household composition so that c_j remains the consumption of the household (Yum, 2022). Risk aversion is captured by σ.

In addition, we assume that parents are biased toward their own educational background. When deciding on their child's school track, they incur a utility cost $\chi(E, S')$ that depends on their own education attainment E and their child's school track S'.[15] In particular, we assume that $\chi(E, S') = \chi_1 \mathbb{I}\{E = 1 \wedge S' = A\} + \chi_0 \mathbb{I}\{E = 0 \wedge S' = V\}$, so that the bias of college-educated parents is governed by χ_1 and that of vocational parents by χ_0.

Moreover, when individuals reach adulthood ($j = 5$), those who opt for and graduate

[14] This introduces non-linear teacher effects. Duflo et al. (2011) empirically provides evidence of those.

[15] We can view this as a result of unmodeled informational friction.

from college face the following utility cost $\psi_1(S, \theta_5, \nu(E^p))$:[16]

$$\psi_1(S, \theta_5, \nu(E^p)) = \exp(\psi + \psi_{S=V} + \psi_\theta \theta_5 + \nu(E^p))$$
$$\nu(E^p) \sim \mathcal{N}(\mu_{\nu,E^p}, \sigma_\nu^2).$$

This cost depends on their secondary school track S, their end-of-school skill level θ_5, and an idiosyncratic "college taste" shock, $\nu(E^p) \sim \mathcal{N}(\mu_{\nu,E^p}, \sigma_\nu^2)$, which is influenced by their parent's education level E^p. This formulation represents two salient features of the transition between secondary and college education in the data that we ask our model to replicate. Firstly, the share of children with an academic track secondary school degree who end up getting a college degree is significantly higher than those with a vocational secondary school degree, which is why the shock depends explicitly on the school track S through $\psi_{S=V}$.[17] Secondly, independently of the school track, the likelihood of college education in the data is increasing in the end-of-school skills, so we allow the college cost to depend on the end-of-school skill level as well through ψ_θ.[18] Finally, to account for additional heterogeneity in the college decision, we allow for normally distributed taste shocks $\nu(E^p)$ that depend on parental education. Its purpose is to reflect heterogeneity in the higher education decision coming from parental background or from channels that are outside of this model.[19]

Finally, we assume that parents are altruistic. Parents take into account the continuation value of their child to a factor of δ when making inter-vivos transfers, thus exhibiting altruism.

[16] The purpose of this cost is to account for the fact that it is, in principle, possible to obtain a college education even after not graduating from an academic track secondary school. However, college education through such "second-chance" opportunities are much less frequent. In Germany, every graduate from an academic track secondary school gets an official qualification that allows for access to academic higher education institutions, while graduates from vocational tracks do not. To go to college, these must either get a qualification through "evening schools", or may be allowed access to certain university degrees after having obtained a higher vocational degree or after having worked for a certain number of years.

[17] In Germany, most of this is coming from the fact that an academic track secondary school diploma automatically qualifies for university entrance.

[18] Net of the above-explained effect coming through the secondary school track graduation, this may partly be due to the fact that for many university degrees, admission is competitive and often even requires a specific end-of-school grade average ("numerus clausus"). Of course, it could also simply reflect the selection of higher-skilled school graduates into an academic career, where these (mostly cognitive) skills are more useful.

[19] Indeed, we recognize that there are many pathways into academic higher education in Germany that fall outside the scope of our model. For example, in some vocational professions, it is possible to become eligible for university entrance after having worked a number of years in that profession. Moreover, the college decision seems to be influenced by parental education, even net of effects going through secondary school track decisions and end-of-school skills. These types of "psychic costs of higher education" are common in the literature (see e.g. Abbott et al. (2019); Daruich (2022)).

During their working life ($j = 5$ to $j = 16$), human capital grows stochastically,

$$h_{j+1} = \gamma_{j,E}\ h_j\ \varepsilon_{j+1}, \quad \log \epsilon_j \sim \mathcal{N}(0, \sigma_\epsilon^2), \tag{6}$$

and the earnings are given by:

$$y_j = w_E\ h_j\ n_j, \tag{7}$$

with n_j the number of hours worked and h_j the accumulated human capital. We assume that all work during the higher education stage ($j = 5$) is of a vocational nature, and therefore, earnings are $y_5 = w_0\ h_5\ n_5$. Note that all prices, including the wage rate w_E, implicitly depend on the distribution of agents in the economy, which we suppress for notational convenience. The human capital is subject to idiosyncratic market luck shocks ε_{j+1}, which we assume follows an i.i.d. normal distribution in logs, with zero mean and constant variance σ_ε^2, as in Huggett et al. (2011).

In the remainder of her life, the agent receives retirement benefits $\pi_j(h_{17})$, which depend on the last human capital level before retirement.[20] Finally, the value of death is normalized to zero.

In the following, we provide the recursive formulation of the agent's decisions. We begin by discussing the decision problem of the working adult with a dependent child, as this is the time at which the agent chooses the school track for her child, which is at the heart of our model. Following this, we will discuss the decision problems of the working adult without dependent children and of the retiree.

3.3 The Optimization Problems

The timing of the model is as follows. At the beginning of each adulthood period, individuals learn about their productivity shock and, in case they have a child, about the child skill shock. Based on this information, an individual decides her consumption (c_j), her savings for the next period (a_{j+1}), and if she is not retired, her hours worked (n_j). We assume that individuals face a life-cycle borrowing constraint

$$a_{j+1} \geq \underline{a}. \tag{8}$$

[20] As is common in the literature, we let benefits depend on human capital in this way to proxy for lifetime earnings, which form the basis of pension benefits in reality.

In addition, an adult must decide on her college education in period $j = 5$, on the inter-vivos transfer in period $j = 13$, and, importantly, on the school track of her child in period $j = 11$.

3.3.1 Work Life and Parenthood, $j = 9, 10, 11, 12, 13$

For each period of the work life, the adult's choices are subject to the human capital production technology in (6), the earnings in (7), a budget and a time constraints:

$$c_j + a_{j+1} = y_j + (1+r)a_j - T(y_j, a_j), \quad c_j > 0, \quad n_j \in [0,1], \tag{9}$$

and the borrowing constraint (8). The interest rate is r, and taxes net of transfers is $T(y_j, a_j)$. Thus, in each period, the individual will optimally allocate her unit of time between hours worked n_j and leisure. She will also decide how to optimally spend her earnings y_j, capital gains, and taxes net of transfers $T(y_j, a_j)$ between consumption c_j and savings a_{j+1}, subject to the life-cycle borrowing constraint $\underline{a}$. For simplicity, we suppress those constraints in the following formulations.

Before The School Track Decision ($j = 9,\ 10$) The child's learning ability ϕ' is realized, and, given the initial skill level is given by (2), the child's skill $\theta'_{j'}$ evolves according to equation (3). A parent with a dependent child in the household solves the following life-cycle savings problem:

$$V_j(E, h_j, a_j, \phi', \theta'_{j'}) = \max_{c_j, a_{j+1}, n_j} \left\{ u(\frac{c_j}{q_j}, n_j) + \beta \int V_{j+1}(E, h_{j+1}, a_{j+1}, \phi', \theta'_{j'+1}) dF(\varepsilon_{j+1}) dF(\eta'_{j'+1}) \right\}$$
$$\text{s.t. } \theta'_{j'+1} = g_{j'}(\theta'_{j'}, E, \bar{\theta'}_{j'}) \ , \quad \eta'_{j'+1} \sim \mathcal{N}(0, \sigma^2_{\eta, j'+1}).$$

The state space includes the parent's and the child's variables. The first state, E, denotes whether the individual is vocational or college educated, h_j her level of human capital determined in the previous period, and a_j, and her savings. In addition, her state space comprises the child's learning ability ϕ' and current skills θ'_j. In the preschool period $j = 1$, the schooling component of the technology of skill formation is zero, and the evaluation of the child's skills does not depend on other children's skills. In the primary school period $j = 2$, the child goes to a comprehensive school, and the evolution of her skills depends on the average skill level of all children in that cohort $\bar{\theta'}_{j'}$.

The School Track Decision ($j = 11$) The school track choice happens at the beginning of secondary school. After the parent observes the realization of her child's skills, θ'_3, she makes the decision whether to send her child to the vocational or academic track school, $S' \in \{V, A\}$. This decision is affected by the value of placing her child in each track, V_{11}, but also by a fixed preference shifter, $\chi(E, S')$, that depends on the child's but also on the parent's educational attainment. Then the value at the beginning of period 11 is given by

$$V_{11}(E, h_{11}, a_{11}, \phi', \theta_3') = \max_{S' \in \{V,A\}} \{W_{11}(E, h_{11}, a_{11}, \phi', \theta_3', S') - \chi(E, S')\}, \tag{10}$$

where the values of sending a child to a school that belongs to school track S' are given by

$$W_{11}(E, h_{11}, a_{11}, \phi', \theta_3', S') = \max_{c_{11}, a_{12}, n_{11}} \left\{ u(\frac{c_{11}}{q_{11}}, n_{11}) + \beta \int V_{12}(E, h_{12}, a_{12}, \phi', \theta_4', S') dF(\varepsilon_{12}) dF(\eta_4') \right\}$$
$$\text{s.t. } \theta_4' = g_3(\theta_3', E, \bar{\theta}'_{3,S'}), \quad \eta_4' \sim \mathcal{N}(0, \sigma^2_{\eta,4}). \tag{11}$$

Future child skills are now affected by $\bar{\theta}_{3,S'}$, which are the average skill levels among children in school track S'. Parents need to form expectations over average skill levels in each track, which in equilibrium, must coincide with the actual distributions. As with prices, including the interest rate on savings r, and the wage rate w_E, we keep the dependence of average skill levels on the aggregate distribution implicit.

Remaining Parenthood ($j = 12, 13$) In period $j = 12$, the parent solves the following problem:

$$W_{12}(E, h_{12}, a_{12}, \phi', \theta_4', S') = \max_{c_{12}, a_{13}, n_{12}} \left\{ u(\frac{c_{12}}{q_{12}}, n_{12}) + \beta \int V_{13}(E, h_{13}, a_{13}, \phi', \theta_5', S') dF(\varepsilon_{13}) dF(\eta_5') \right\}$$
$$\text{s.t. } \theta_5' = g_4(\theta_4', E, \bar{\theta}'_{4,S'}), \quad \eta_5' \sim \mathcal{N}(0, \sigma^2_{\eta,5}), \tag{12}$$

where the child's school track S' that has been decided in the previous period is now included in the parent's state space.[21]

Just before her child reaches the age of 18 and becomes independent, the parent decides on a financial inter-vivos transfer that her child receives, a_5, while taking into account the

[21] That is, we assume that there is no track-switching possibility during secondary school. For example, in 2010/11, only around 2.5% of children in the first stage of secondary school in Germany switched school tracks (Bellenberg and Forell, 2012). Moreover, this number includes switches among different tracks that we group into the vocational track, so is likely an upper bound of the track switches between the vocational and academic tracks. The great majority of track switches are from an academic track school to a vocational track school rather than the other way.

child's future value V_5'. As in Daruich (2022), we model this as an interim decision problem and assume that the parent already knows the realization of her own market luck shock and her child's final skill shock but does not know the realization of the college taste shock $\nu'(E)$. As is common, the transfer cannot be negative, so parents cannot borrow against the future income of their child. The value at the beginning of period 13 is then

$$
\begin{aligned}
V_{13}(E, h_{13}, a_{13}, \phi', \theta_5', S') = \max_{a_5' \geq 0} & \left\{ \widetilde{V}_{13}(E, h_{13}, a_{13} - a_5') + \delta\, \mathbb{E}_\nu\, V_5'(\theta_5', a_5', \phi', S', E) \right\} \\
\nu'(E) \sim \mathcal{N}(\mu_{\nu,E}, \sigma_\nu^2), &
\end{aligned}
\tag{13}
$$

where $\widetilde{V}_{13}$ is the value for a parent with savings a_{13} after the inter-vivos transfer has been made

$$
\begin{aligned}
\widetilde{V}_{13}(E, h_{13}, a_{13}) &= \max_{c_{13}, s_{14}, n_{13}} \left\{ u(c_{13}, n_{13}) + \beta \int V_{14}(E, h_{14}, a_{14}) dF(\varepsilon_{14}) \right\} \\
\text{s.t. } c_{13} + a_{14} + a_5' &= y_{13} + (1+r)a_{13} - T(y_{13}, a_{13}).
\end{aligned}
\tag{14}
$$

where the transfer a_5' enters the budget constraint.

3.3.2 Work Life Without a Dependent Child, $j = 5, 6, 7, 8$ and $j = 14, 15, 16$

As before, for each period of the work life, the choices of the working adult's without a dependent child are subject to the human capital production technology in (6), the earnings in (7), the budget and the time constraints in (9) and the borrowing constraint (8). For simplicity, we also suppress them in the following formulations.

Independence ($j = 5$) At the beginning of adulthood ($j = 5$), the human capital h_5 depends on past child skills following (4). The newly independent adult solves the following problem:

$$
\begin{aligned}
V_5(\theta_5, a_5, \phi, S, E^p) = \max\{ & W_5(E = 0, h_5, a_5, \phi), \\
& W_5(E = 1, h_5, a_5, \phi) - \psi_1(S, \theta_5, \nu(E^p))\} \\
\nu \sim \mathcal{N}(\mu_{\nu,E^p}, & \sigma_\nu^2),
\end{aligned}
\tag{15}
$$

where W_5 denotes the values of the college and non-college education, given by

$$
W_5(E, h_5, a_5, \phi) = \max_{\substack{c_5, a_6 \\ n_5 \in [0, \bar{n}(E)]}} \left\{ u(c_5, n_5) + \beta \int V_6(E, h_6, a_6, \phi) dF(\varepsilon_6) \right\}.
\tag{16}
$$

In addition to choosing her consumption level, savings, and hours worked the newly independent adult chooses whether to graduate from college or not. Her state space includes her past child skills θ_5, her savings that she received from her parent a_5, and her learning ability that she will eventually (imperfectly) transfer to her child. The last two state variables are her school track S and the higher education state of her parent $E^p \in \{0,1\}$, which influence her utility cost of going to college $\psi(S, \theta_5, \nu(E^p))$. Note that during her college education, an individual may also work, but only at the vocational wage rate w_0. However, obtaining a college education reduces the time available for work, as individuals spend part of their total time endowment studying, thus $\bar{n}(E=1) < 1$.[22]

Remaining Work Life ($6, 7, 8$ and $j = 14, 15, 16$) For the two following periods $j = 6, 7$, an adult without a child solves the following life-cycle savings problem:

$$V_j(E, h_j, a_j, \phi) = \max_{c_j, a_{j+1}, n_j} \left\{ u(c_j, n_j) + \beta \int V_{j+1}(E, h_{j+1}, a_{j+1}, \phi) dF(\varepsilon_{j+1}) \right\}. \tag{17}$$

The state space only includes the parent's related variables, including her learning ability that will be transferred (imperfectly) to her child.

In period $j = 8$, the individuals know that they will become parents next period. For that reason, they take expectations over the learning ability of their future child, ϕ', on top of the expectations over the market luck shocks. We assume that ability is imperfectly transmitted from parents to children, according to $\phi^c \sim G(\phi'|\phi)$ and (1). Thus, the value in period 8 becomes

$$\begin{aligned} V_8(E, h_8, a_8, \phi) &= \max_{c_8, a_9, n_9} \left\{ u(c_8, n_8) + \beta \int V_9(E, h_9, a_9, \phi') dF(\varepsilon_9) dG(\phi'|\phi) \right\} \\ \text{s.t. } & \log \phi' = \rho_\phi \log \phi + \epsilon_\phi, \qquad \epsilon_\phi \sim \mathcal{N}(0, \sigma_\phi^2). \end{aligned} \tag{18}$$

For the rest of the periods $j = 14, 15, 16$, an adult whose child became independent solves the following life-cycle savings problem:

$$V_j(E, h_j, a_j) = \max_{c_j, a_{j+1}, n_j} \left\{ u(c_j, n_j) + \beta \int V_{j+1}(E, h_{j+1}, a_{j+1}) dF(\varepsilon_{j+1}) \right\}, \tag{19}$$

where the learning ability ϕ has already been transmitted to the child and does not enter the state space anymore.

[22]Note that, compared to studies that focus on the U.S. we do not model college costs as monetary costs. This is because most colleges in Germany are public and have very low tuition fees.

3.3.3 Retirement, $j = 17, 18, 19, 20$

Everybody retires at the beginning of model period 17, which corresponds to age 66 in real life, and receives retirement benefits $\pi_j(h_{17})$. After period 20, that is, at age 82, agents die with certainty and exit the model. The values in these periods are

$$\begin{aligned} V_j(E, h_{17}, a_j) &= \max_{c_j > 0, a_{j+1} \geq \underline{a}} \{u(c_j, 0) + \beta V(E, h_{17}, a_{j+1})\} \\ \text{s.t. } c_j + a_{j+1} &= \pi_j(h_{17}) + (1+r)a_j - T(0, a_j). \end{aligned} \tag{20}$$

3.4 Aggregate Production, and Government

We assume that a representative firm produces output according to the Cobb-Douglas production function $Y = K^{\alpha} H^{1-\alpha}$, where K is the aggregate physical capital stock and H is a CES aggregate of total labor supply, which is defined by:

$$H = [\omega H_0^{\epsilon} + (1-\omega) H_1^{\epsilon}]^{\frac{1}{\epsilon}}. \tag{21}$$

Here, H_0 is the aggregate labor supply in efficiency units of workers with vocational higher education, and H_1 is that of workers with a college education. The physical capital stock depreciates at rate δ_f.

The government taxes labor income progressively, such that labor income net of taxes amounts to $y_{net} = \lambda y^{1-\tau_n}$ (Heathcote et al., 2017). It also taxes capital income linearly according to $\tau_a r a_j$ (Yum, 2022). All tax revenue is used to finance retirement benefits π_j as well as fixed lump-sum social welfare benefits g that are paid to every household. These may include child allowances, unemployment benefits, or contributions to health insurance.

3.5 Equilibrium

We solve for the model's stationary equilibrium and its associated distribution using the numerical strategy in Lee and Seshadri (2019). Stationarity implies that the cross-sectional distribution over all states in every period j is constant across cohorts. Our model economy consists of 20 overlapping generations or cohorts at each time. The equilibrium requires that households and firms make optimal choices according to their value functions and firm first-order conditions, respectively. Moreover, the aggregate prices for physical capital and both types of human capital r, w_0, and w_1 are competitively determined and move to clear all markets. Note that we do not require the government budget to clear as well. Instead, we assume that all government revenues that exceed the financing of all social welfare programs result in linearly independent spending.

A special feature of our model is that learning during the school years depends on the distribution of children across school tracks. Importantly, an equilibrium requires that parents form expectations over average skill levels in each track, which in equilibrium, coincide with the actual distributions. A detailed definition of the equilibrium is given in Appendix A.2.

4 Developing Intuition: Child Skill Formation and School Tracking

In order to develop an intuition for the key mechanisms at work for the timing of school tracking, we develop a parsimonious economic model of the formation of a child's skills during her school years, which we think of as encompassing ages 6 to 18 (period 2 to 4).

4.1 Child Skill Formation

Suppose each child i arrives just before entering school with a set of skills that can be summarized in a univariate level, $\theta_{i,j=1}$.[23] Note that recognizing the multistage nature of child development is crucial not only to capture the self-productive nature of skills and dynamic complementarity of investments into skills but to differentiate the effects of school tracking at different stages of the schooling years. In the remaining, we simplify the notation by omitting the subscript i. We specify the following form for the child skill technology, where θ_j refers to the child i's skill level in period j:

$$\begin{aligned}\theta_{j+1} &= \theta_j + \alpha\bar{\theta}_{-i,j} + \beta p_j - \frac{\gamma}{2}\left(\theta_j - p_j\right)^2 + \frac{\gamma}{2}\theta_j^2 + \zeta E + \eta_{j+1} \\ &= \theta_j + \alpha\bar{\theta}_{-i,j} + \beta p_j - \frac{\gamma}{2}{p_j}^2 + \gamma\theta_j p_j + \zeta E + \eta_{j+1},\end{aligned} \tag{22}$$

where η_{j+1} describes unobserved shocks to the formation of skills, that are assumed to be independent and normally distributed around mean zero with a variance $\sigma^2_{\eta_{j+1}}$. We assume

[23] As in Cunha and Heckman (2007), we do not differentiate between abilities and skills, as both are partly endogenously produced and partly exogenously determined pre-birth. Moreover, we do not allow for potentially different production technologies of cognitive and non-cognitive skills as in Cunha et al. (2010) or Daruich (2022). Instead, in the tradition of Becker and Tomes (1986), we focus on one composite skill, which after school can be translated into adult human capital that is rewarded on the labor market.

the parameters $\gamma > 0$ and $\alpha,\ \beta,\ \zeta \geq 0$. Parental background affects the child's skill evolution linearly through ζE. In addition, the evolution of the child's skills is directly affected by past skills and through interactions with peers $\bar{\theta}_{-i,j}$, similar to Duflo et al. (2011).[24] Finally, the pace of instruction is denoted by p_j. Learning monotonically increases with the pace p_j at rate β and decreases with the distance between a child's current skill and the pace of instruction. It implies that for each skill level θ_j, there exists a unique optimal pace of instruction p_j^* that maximizes the child's future skill level, keeping everything else fixed. The optimal pace of instruction p_j^* for a child with skills θ_j is given by $p_j^* = \frac{\beta}{\gamma} + \theta_j$.[25] It is strictly increasing in current skill, such that higher-skilled children also prefer a higher pace of instruction. With β reasonably small, these assumptions imply that for a child with a very low current skill level attending a class with a very ambitious, high instruction pace can be harmful to the point when she actually loses skills. Similarly, a high-skilled child might be so sub-challenged in a class with a very low pace that she actually loses skills.

4.2 Optimal Pace of Instruction

Consider a policymaker choosing the pace of instruction, which we think as reflecting the curriculum as well as the teaching intensity, for a group of children in a given school track S. Let us assume the policymaker seeks to maximize the aggregate next-period skills of this group of children and knows the child skill formation in (22).

Lemma 1. *The optimal pace of instruction, given a distribution of child skills in a given track S, is given by*

$$P_{j,S}^* = \frac{\beta}{\gamma} + \bar{\theta}_{j,S} \tag{23}$$

Proof. Follows from taking the first order condition of the conditional expected value $\mathbb{E}(\theta_{j+1|S})$ in (22) with respect to p_j under the i.i.d. assumption of η_{j+1}. □

Lemma 1 establishes that the optimal pace of instruction a policymaker would pick for the group of children in track S will be a function of the average skill level of this group, $\bar{\theta}_{j,S}$. It is thus only (first-best) optimal for a child with exactly the average skill level in that track. Every child with a skill level above or below the average loses in terms of future skills

[24] We concentrate on the case with a linear-only direct peer externality, governed by α. As summarized in Epple and Romano (2011), many studies find that such linear-in-means peer effects are sizable and robust across settings. Evidence on non-linear peer effects in the classroom is more ambiguous. For that reason, we do not incorporate non-linearities in peer effects directly. Instead, we consider the endogenous setting of instruction levels across school tracks as a channel through which non-linear peer effects arise.

[25] This follows from taking the first order condition of the child skill formation in (22) with respect to p_j^*.

compared to a world in which she would be taught at her individually optimal level.[26] Note that replacing the pace of instruction in (22) by the optimal one gives the model child skill formation (3) where $\kappa_{0,j} = \frac{\beta^2}{2\gamma}$, $\kappa_{1,j} = (1 + \beta)$, $\kappa_{2,j} = -\kappa_{4,j}$, and $\kappa_{3,j} = \alpha$, $\kappa_{4,j} = \frac{\gamma}{2}$ and $\kappa_{5,j} = \zeta$.

4.3 School Tracking

Let the schooling system in a given stage be characterized by the number of distinct school tracks S. If there is only one track to which all schools belong, we speak of a *comprehensive system*. If there are two distinct school tracks, we speak of a *tracking system*.[27] Moreover, we assume that once a comprehensive system is switched to a tracking system between school stages, it cannot switch back to a comprehensive system. Finally, we do not allow children to switch between school tracks.[28] The pace of instruction in each track is chosen by a policymaker seeking to maximize aggregate end-of-school skills as in (23).[29] Finally, note that we assume the heterogeneity in instruction paces across tracks does not entail systematic differences in teacher quality or resources devoted to teaching across tracks that could also affect child skill development.[30]

In the following, we discuss the sorting mechanism of children across tracks and compare the distribution of the end-of-school skills between both systems, assuming away direct parental inputs ($\zeta = 0$). We start by considering the case with one period of schooling only,

[26]Clearly, in a first-best world, a policymaker would like to provide every child with her preferred pace of instruction, which would trivially maximize end-of-school skills.

[27]Thus, while in principle, a large number of school tracks is conceivable, we restrict a tracking system to two school tracks as this corresponds to a typical number across OECD countries.

[28]To the best of our knowledge, there are no cases in OECD countries, where a comprehensive system follows a tracking system. In virtually all countries, the schooling years start with comprehensive primary school. Tracking into distinct school types then occurs, if at all, at some point during secondary school. Among OECD countries, the first age of school tracking varies from age 10 in Austria and Germany, to age 16 in Australia, Canada, Chile, Denmark, Finland, Iceland, New Zealand, Norway, Poland, Spain, Sweden, the United Kingdom, and the United States (OECD, 2013). While, in principle, switches between tracks during the secondary school years are possible, they are very rare. Dustmann et al. (2017) document a share around 2% of track switches during secondary school in Germany.

[29]For example, in Germany, the curricula in the different tracks are set by each federal state under some general federal education goals. They consist of learning and competence goals as well as methods and specific topics that should be taught, separately for each school track, subject, and school grade. The curricula are subject to frequent review and renewal. For example, as of 2023, 14 out of 16 federal states in Germany updated the curriculum in the last four years and 7 out of 16 in the last two years.

[30]The literature on international differences in student achievement tends to find limited effects of resources spent per student on learning outcomes (Woessmann, 2016). In Appendix A.3, we summarize information on expenditure per student as well as teacher quality across different school tracks in Germany. While we do not necessarily abstract from these factors in affecting child skill development, we conclude that they are not correlated with school track.

so that θ_2 are the skills at the end of school. This means that either a tracking system or a comprehensive system can be implemented for the entirety of the school year.

4.3.1 One-period Schooling System

Sorting Mechanisms We consider two alternative allocation mechanisms. In the first one, a policymaker who seeks to maximize the aggregate end-of-school skills allocates children across tracks. The second alternative consists of parents making the track decision for their child, with the goal of maximizing their child's end-of-school skill level.[31] We assume that parents know the skill formation technology (22) as well as the distribution of skill levels among all children and the skill shock distribution. Moreover, it is common knowledge that policymakers set the pace of instruction according to (23). In equilibrium, expectations about the paces of instruction are consistent with the actual paces of instruction.

Proposition 1 shows that, in both alternatives, the track decision is governed by a sharp cut-off skill level. A policymaker would optimally split the distribution exactly at its median.[32] Intuitively, this generates the highest aggregate end-of-school skills as it minimizes the variance of skills in each track, or in other words, it creates peer groups that are as homogeneous as possible. In doing so, the policymaker internalizes that any effects coming from the direct peer externality exactly offset each other across tracks. Thus, all gains achieved from making average peer skills in one track higher are lost as the average level in the other track becomes smaller.

In contrast, if parents endogenously sort their children into the two tracks, they make their decision irrespective of the aggregate outcomes. Whenever the direct peer effects are positive ($\alpha > 0$), the cut-off skill level is smaller than the optimal threshold a policymaker would pick. More children would go to the highest peer group track. Intuitively, parents do not internalize the effect their child's skills have on aggregate end-of-school skills. The threshold is characterized by the skill level $\tilde{\theta}_1^*$ at which the child's next period skill is in expectation, the same in both tracks.

Proposition 1. *In Equilibrium, the allocation of children across tracks is characterized by*

[31] This has become common practice in Germany, where in the majority of federal states, parents are completely free in making the secondary school track choice for their children. Only in three states, Bavaria, Thuringia, and Brandenburg, academic school track access is conditional on a recommendation by the primary school teachers. These recommendations are often tied to achieving a certain grade point average in math and German in primary school. However, even in these states, children without a recommendation can take advantage of a trial period in an academic track school, after which the child will be assessed again.

[32] Equivalent to the mean in this context. A similar argument has been made repeatedly in the theoretical literature. See for instance, Epple and Romano (2011).

a skill threshold $\tilde{\theta}_1$, such that all children with initial skills below $\tilde{\theta}_1$ go to one track and all children with initials skills above $\tilde{\theta}_1$ go the other track.

- *If the track allocation is done by the policymaker, the optimal skill threshold corresponds to the average initial skill level $\tilde{\theta}_1^* = \mathbb{E}[\theta_1] = 0$.*

- *If the track allocation is done by parents, the endogenous skill threshold that emerges from this game depends on the direct peer externality α. With $\alpha > 0$, the threshold is smaller than $\tilde{\theta}_1^*$.*[33]

Proof. In Appendix A.1. □

The End-of-school Distribution Proposition 2 describes the end-of-school distribution of skills in both schooling systems in the one-period model. Independently of the sorting mechanism, the average end-of-school skills in a full optimal tracking system are always larger than in a comprehensive system. Intuitively, this advantage comes from more homogeneous peer groups in each track, in terms of their initial skills. Since learning generally decreases in the variance of skills among children in a track, more homogeneity on average increases end-of-school skills.

A full tracking system necessarily leaves a non-negative mass of children worse off compared to a comprehensive system. These children have initial skills around the tracking threshold and would be closer to their optimal instruction pace in a comprehensive system. In an optimal tracking system with $\tilde{\theta}_1 = 0$, these children thus occupy the center of the distribution and would, given a choice, prefer a comprehensive system.[34] If there are no direct peer effects, an equal share of children in both tracks lose relative to the comprehensive counterpart. However, with positive peer effects the losses are concentrated among the track with the lower average peer level. This reflects a robust finding of the empirical school tracking literature that especially the children at the bottom of the skill distribution suffer from a tracking system (see, e.g., Matthewes (2021)).

Proposition 2.

- *Aggregate end-of-school skills in a full tracking system are larger than in a full comprehensive system. This holds regardless of who makes the track decision, i.e. regardless of the tracking skill threshold. $\tilde{\theta}_1$*

[33] We rule out an (uninteresting) equilibrium of the track choice game in which parents randomly allocate their child into one of the two tracks, leading to the same distribution of skills in both tracks and, hence, the same pace of instruction.

[34] This is interesting in a political economy context as the median voter in this model would prefer a comprehensive system. This could partially explain why we see different tracking systems across different countries.

- *Children with initial skills inside a non-empty interval lose from a full tracking system in terms of their end-of-school skills relative to a fully comprehensive system. With $\alpha = 0$, the losses are symmetric in both tracks. With $\alpha > 0$, the losses are concentrated in the track with the lower average skill level.*

Proof. In Appendix A.1. □

Note that these results are not affected by the presence of skill shocks in the one-period model. This is because these shocks are assumed to be mean zero and realized at the end of the period.[35] They matter in a two-period schooling system.

4.3.2 Two-period Schooling System

We consider a two-period schooling system, in which the pace of instruction is defined by (23) in each of the two periods and where θ_3 represent end-of-school skills that are maximized in expectation. We are interested in a comparison between the end-of-school skill distribution in an early tracking system, ET, and a late tracking system, LT in which the allocation is done by the policymaker. The early tracking system is characterized by an initial track allocation into two tracks, V and A, at the beginning of the school year, $j = 1$.[36] In an early tracking system, a child, therefore, remains in her school track for the two school periods.[37] The late tracking system is characterized by all children going to a comprehensive school in the first period, followed by the allocation into schools that belong to one of two tracks, V and A, at the beginning of the second period. Importantly, while in the LT case, this allocation occurs after the skill shock η_2 is realized, in the ET case, the allocation occurs before.[38]

Proposition 3 shows that aggregate end-of-school skills in an optimal LT system can be larger than in an optimal ET system if the variance of the skill shocks is large enough. Intuitively, in the presence of skill uncertainty, early tracking generates misallocation that leads to learning losses not only for the misallocated individual child but also poses an externality

[35] Therefore, the realization of shocks affects the end-of-school distribution only in that it raises the variance uniformly in both tracks.

[36] As argued before, we do not allow for track switches during the schooling years, including at the beginning of $j = 2$.

[37] Absent any costs of re-tracking, a school system that features such a re-tracking possibility would improve aggregate end-of-school skills in our model relative to the early tracking system we describe here. However, we focus on the early tracking system, as in reality, re-tracking (or second chance) opportunities *during* school years are used only relatively rarely.

[38] We do not consider a fully comprehensive system in which children remain in comprehensive track schools for the whole duration of their school career. Proposition 2 implies that such a system cannot achieve higher aggregate end-of-school skills compared to a late tracking system.

for all other children as the instruction pace is endogenous to the peer composition. Indeed, in the second period, the unexpected skill shocks render the peer group more heterogeneous. Depending on the size of the skill uncertainty, the efficiency gains of the ET system in the first period can be outweighed by the losses due to misallocation in the second period.

Proposition 3. *Average end-of-school skills in the two-period model are larger in an optimal late tracking system than in an early tracking system iff*

$$\frac{\sigma^2_{\eta_2}}{\sigma^2_{\theta_1}} > 1 + \alpha + \alpha^2 + \beta + \frac{\beta^2}{2} + 2\alpha(1+\beta) + \frac{\gamma^2}{2\pi}\sigma^2_{\theta_1}. \tag{24}$$

Proof. In Appendix A.1. □

This parsimonious modeling provides a reasoning for the child skill formation (3) and helps see the main mechanisms at work for school tracking: efficiency gains of homogenous peer groups and misalocation losses due to the skill uncertainty. The efficiency of the timing of a tracking system will then not only depend on estimates of the (age-specific) child skill formation technology parameters but also crucially on the size of the skill uncertainty. However, to obtain analytical results and form intuition, we ignored important features of the full model that also influence the efficiency of school tracking policies. In particular, we have defined simple objective functions, namely the maximization of (aggregate) end-of-school skills.[39] Still, in a richer model, parents and policymakers can have different objectives. For example, the policymaker could maximize the aggregate output in (21), and parents would consider future labor market prospects and may be biased toward their own educational path. This bias is another source of misalocation to be considered when evaluating the effects of the timing of school tracking. In addition, even though graduation from an academic track school often leads to college graduation, it is not systematic. Moreover, the wage rate is a function of labor demand by firms that likely employ labor from both vocational and academic degrees. Thus, general equilibrium effects, as well as later college choices, are also important to consider to provide insights into the effects of the timing of school tracking policies on lifetime and intergenerational outcomes.[40] To take into account all those factors

[39]We have focused the debate on aggregate end-of-school skills as we assume this is the main goal of a policymaker. One could also think of different objectives, though. For example, a policymaker could take into account the inequality in the child skill distribution or could seek to guarantee a minimum level of skills for every child. In such cases, the conclusions about the implications of tracking but also regarding the optimal instruction pace-setting strategy are likely to change.

[40]See Dustmann et al. (2017) for an excellent discussion about the pathways through the German Education System. In particular, it is the period between the end of secondary school and the beginning of possible tertiary education that is characterized by multiple possibilities. In fact, in their paper, these second-chance

in the evaluation of the effects of school tracking policies, we now calibrate a quantitative version of the full model.

5 Model Calibration

As is common in the literature, we parameterize the model following a two-step approach. In the first step, we estimate the parameters of the child skill formation technology during the school years, as well as other selected model parameters directly from the data. In the second step, the remaining parameters are estimated using the simulated method of moments by matching the moments from the stationary equilibrium distribution of the model to their empirical counterparts. A summary of the externally calibrated parameters is given in Table 6 and of the internally estimated ones in Table 7.

5.1 Data and Sample Selection

All externally estimated parameters in the first step and moments used in the second step are based on two data sources. The first source is the German National Educational Panel Study (NEPS) described in Section 2. We further restrict the sample to individual observations containing information on the school and class in that school a child attended in a given year.

The second data source is the German Socio-Economic Panel (SOEP), an annual representative survey from which we use the 2010-2018 waves. The data contains rich information on labor supply, income, and education on the individual level. We use this data source primarily to construct empirical moments for the working stage of the life cycle, as will be detailed below. For this reason, the only sample selection that we do is dropping those with hourly wages below the first and above the 99th percentile. Lastly, we convert all income data to 2015 Euros using a CPI index for inflation adjustment.

We begin by detailing how we measure, identify, and estimate the parameters of the child skill formation technology, as these constitute the most important ingredient of our model. Then, we describe the functional forms and estimation strategies for all remaining parameters.

opportunities are the main reason why the initial track choice does not have an impact on the marginal child.

5.2 Estimation of the the Child Skill Formation Technology

We specify the production technology of (the logarithm) of child i's skills (3) that we take to the data as follows:[41]

$$\theta_{i,j+1} = \kappa_{0,j} + \kappa_{1,j}\theta_{i,j} + \kappa_{2,j}{\theta_{i,j}}^2 + \kappa_{3,j}\bar{\theta}_{-i,j,S} + \kappa_{4,j}{(\theta_{i,j} - \bar{\theta}_{j,S})}^2 + \kappa_{5,j}E_i + \eta_{i,j+1}, \tag{25}$$

Note that (25) is just a rearranged version of the child skill technology (22) after substituting in the optimal pace of instruction. We denote by $\bar{\theta}_{-i,j,S}$ the average skill level of child i's *classroom* peers, as opposed to $\bar{\theta}_{j,S}$, which refers to the average skill level of all children in a school that belongs to track *theta* and arises from optimal track-specific instruction paces. Note that in the model, $\bar{\theta}_{-i,j,S} = \bar{\theta}_{j,S}$, since we assume a representative school and class per track (or alternatively, identical classes conditional on school tracks). In the data, however, there is clearly heterogeneity across classes, even within a school track. Since we are interested in capturing skill development effects that arise from direct interactions with peers, which are likely occurring in a specific classroom, we, therefore, exploit that heterogeneity in the estimation. During primary school, we observe only one comprehensive track in the data. In that case, even with classroom-specific direct peer effects, we cannot fully identify the parameters in (25), which is why we drop θ^2.[42]

In the estimation, the parental educational attainment E is a time-constant dummy that equals 1 if child i comes from a household in which at least one parent is college educated. We use test scores to measure the evolution of this skill measure.[43] As is common in the child skill formation literature (Cunha et al., 2010; Agostinelli and Wiswall, 2016), we think of log child skills θ as latent variables that are only imperfectly measured in the data. For that reason, we employ a linear measurement system for the logarithm of latent skills in each period and identify the loadings on each measure in each period by ratios of covariances of the measures by subject (as in Agostinelli et al. (2019)).[44]

[41] Following the work in Cunha et al. (2010), much of the empirical and quantitative literature using child skill formation technologies employed parametric specifications of (3) of the constant elasticity of substitution (CES) form. As noted in Agostinelli and Wiswall (2016), this requires, under standard parameter restrictions, that all input factors are static complements. An alternative is to use a nested CES structure as in Fuchs-Schündeln et al. (2023); Daruich (2022). To retain tractability, we follow Agostinelli and Wiswall (2016) and opt for the trans-log approach.

[42] This is also the reason why we prefer (25) over a model that includes $\bar{\theta}^2_{j,S}$ and the interaction $\theta\bar{\theta}_{j,S}$ as separate regressors, such as (22), even when using class-specific peer effects. While in version (25), we just have to drop the squared term on skills, which is typically statistically insignificant even when two tracks are available, in version (22), we cannot identify either the coefficient in front of $\bar{\theta}^2_{j,S}$ or that of $\theta\bar{\theta}_{j,S}$.

[43] As argued in Borghans et al. (2008), achievement test scores measure both cognitive and non-cognitive skills.

[44] Appendix Section provides details on skills measurement. Appendix Tables S3 and S4 describe the

We present our preferred IV estimates of the child skill production technology parameters for all three stages of the schooling career in Table 4.[45] Recall that $\theta_{i,j}$ is defined as the logarithm of child skills. Hence, we can interpret the coefficients as elasticities. Thus, $\hat{\kappa}_{1,2} = 0.954$ means that a 1% increase in latent skills at the beginning of primary school is associated with an 0.954% increase with end-of-primary school skills. Generally speaking, the own-skill elasticity is close to one for the first two stages and decreases in the second half of secondary school, suggesting a relatively high own-skill productivity, as is commonly found in the literature (see estimates in Cunha et al. (2010); Agostinelli et al. (2019)). During secondary school, the estimated coefficient $\hat{\kappa}_2$ are positive. More importantly, we cannot reject the hypothesis that $\hat{\kappa}_2 = -\hat{\kappa}_4$ which is in line with Section 4.

The estimated coefficients $\hat{\kappa}_4$ are always negative and statistically significant at the 10% level. They indicate that a 1% increase in the squared distance to the average skill level in a track is associated with an up to 0.72% decrease in the next period's skills. This lends empirical support to the idea that the instruction pace in every track is tailored to the average skill level, and deviations, in both directions, from this level can hurt individual skill development. The estimated $\hat{\kappa}_3$ are generally small and often statistically insignificant.[46]

The final estimates we use to parameterize the child skill formation technology in our model are the sets $\kappa_{n,j}$ for $n = 0, 1, 2, 4, 5$ and $j = 2, 3, 4$ coming from the IV estimates in Table 4. That is, we do not include a direct peer externality in the model, given the insignificant estimates of $\kappa_{3,j}$. For the first stage of secondary school, $j = 3$, we opt to use the parameter estimates from the age G7-G9 sample.

5.3 Remaining Parameters

5.3.1 Preferences

We set the inverse elasticity of intertemporal substitution, $\sigma = 2$, a value that is common in the literature. The Frisch elasticity of labor supply is set to 0.5. The disutility shifter b is estimated internally in order to match the average time worked in the SOEP data, given that the total time available after sleep and self-care is normalized to 1.

evolution of child skills over time using the identified latent variables.

[45]The baseline OLS estimates in Appendix Table S5. The IV estimates are typically larger than the OLS estimates. This indicates the importance of measurement errors and our correction of it.

[46]In Appendix A.6, we repeat the estimation using the average skill level across all children in a given track rather than in a classroom. The estimated coefficients are positive and slightly larger, indicating that differences across tracks (and hence across instruction paces) are more important than differences across classrooms.

Table 4: IV Estimates using Class-specific direct Peer Effects

		Dependent Variable: $\theta_{i,j+1}$ in model period			
		$j=2$	$j=3$		$j=4$
			Grades		
Coefficient	Variable	G1-G4	G5-G9	G7-G9	G9-G12
$\hat{\kappa}_{1,j}$	$\theta_{i,j}$	0.954	1.001	0.955	0.778
		(0.027)	(0.051)	(0.047)	(0.051)
$\hat{\kappa}_{2,j}$	$\theta^2_{i,j}$	-	0.366	0.396	0.300
		-	(0.210)	(0.202)	(0.143)
$\hat{\kappa}_{3,j}$	$\bar{\theta}_{-i,j,S}$	-0.043	0.283	-0.199	-0.08
		(0.069)	(0.212)	(0.133)	(0.108)
$\hat{\kappa}_{4,j}$	$(\theta_{i,j} - \bar{\theta}_{j,S})^2$	-0.215	-0.559	-0.719	-0.596
		(0.136)	(0.359)	(0.325)	(0.204)
$\hat{\kappa}_{5,j}$	$E=1$	0.022	0.007	0.011	0.008
		(0.007)	(0.009)	(0.008)	(0.007)
$\hat{\kappa}_0$	Constant	-0.043	-0.101	-0.190	-0.072
N Children		3,529	1,934	2,580	2,934
N Schools		326	137	188	240
R^2		0.365	0.410	0.500	0.449

Notes: This table presents the coefficient of regressions of skills on past skills, past skill squared, average peers, distance to the average skill in the track squared, and parent's education dummy. All observations are weighted. Standard errors are clustered at the school level. Models control for age, gender and school-fixed effects. Source: NEPS

We set the time discount factor β, such that the equilibrium interest rate amounts to 4% annually. The altruism parameter δ is calibrated such that the ratio of average inter-vivos transfers to average labor income in the model corresponds to that of average higher education costs of children to average 4-year labor income in the data. According to a survey by the German Student Association in 2016, the monthly costs of living during the higher education stages range from 596 to 1250 Euros (Middendorf et al. (2019)). We expect the parents to bear the bulk of these costs and assume that they support their child for an average of 5 years (the length of time it takes to complete studies that are equivalent to a masters level). Then, the ratio of total costs to average 4-year labor income ranges from 0.32 to 0.67. In our baseline calibration, we take as a target a ratio of 0.6.

Finally, we estimate the bias parameters χ_1 and χ_0 to match the share of deviations from secondary school track recommendations by parental education in the data.

5.3.2 Initial Child Skills and Child Skill Shocks

Initial child skills just before entering primary school are a function of the learning ability of a child, which is imperfectly transmitted from the parent following an AR(1) process with inter-generational correlation coefficient ρ_ϕ, and variance σ^2_ϕ. Since the learning ability is correlated with the eventual higher education outcome of a parent, we pick as the target moment for ρ_ϕ the difference in average preschool child skills by parental education in one standard deviation unit. The variance σ^2_ϕ is then estimated to match the variance of initial math test scores in the data.

An integral part of the child skill development is the presence of unforeseeable, permanent shocks to child skills. As discussed in Section 4, the size of such shocks has important implications for the effects of school tracking policies as they can give rise to efficiency losses if "late-bloomer" effects are large. To quantify the importance of child skill shocks in our model, we internally estimate the shock variance $\sigma^2_{\eta,j+1}$, for $j = 2, 3, 4$. As target moments, we choose the correlation of a child's math test score percentile rank across periods.[47] In this way, we capture all changes in a child's relative position in the skill distribution in a given period that cannot be accounted for by the skill formation technology or track choices.[48]

[47] Appendix Table S4 describes the correlations of child skills between periods using the identified latent variables.

[48] In reality, such changes may also arise from factors that are outside the scope of this model but can put children on a different skill formation path. These could be, for example, a change of schools within a school track, a change of teachers within a class, or even tutoring sessions that are uncorrelated with parental education.

5.3.3 College Costs

We estimate the two parameters ψ and $\psi_{S=V}$ of the college costs to match the share of graduates from an academic secondary school who follow up with a college education and the share of vocational secondary school graduates that end up in college. We discipline the coefficient ψ_θ that multiplies end-of-school skills by matching the regression coefficient on log math test scores from a regression of a college graduation dummy on end-of-school test scores.

The normally distributed college taste shock ν, with parental education specific means μ_{ν,E^p} and variance σ^2_ν, account for additional heterogeneity in the college decision. We calibrate the two parameters $\mu_{\nu,E^p=0}$ and $\mu_{\nu,E^p=1}$ to match the share of children from each parental education background that receive a college degree in the data. Finally, we calibrate the variance, σ^2_ν to match the variance of the residuals from the above regression of college education on end-of-school skills, as in Daruich (2022).

The final component of college costs is not a part of the "psychic" costs ψ_1 but reflects the time cost of obtaining a college education. We assume that studying for a college degree takes away around 60% of the total time available for work for four years or one model period.[49] Thus, we set the maximum remaining time during the higher education stage to $\bar{n}(E = 1) = 0.40$.

5.3.4 Human Capital Growth

We set the child-skill-to-human-capital anchor, ξ, such that in equilibrium average labor income before taxes is equal to 1 (Lee and Seshadri, 2019). We estimate the deterministic human capital growth profiles for both types of education, $\{\gamma_{j,E}\}$, $j = 5, ..., 16$ using wage regressions in the SOEP data, following the approach in Lagakos et al. (2018).[50] The resulting

[49] A common estimate is that full-time studying takes around 40 hours per week, which amounts to around 60% of the maximum weekly work hours, which we set to 65. Moreover, the average study length in Germany is 8 semesters or 4 years.

[50] Concretely we create, separately for each education group, 4-year work experience bins. We then estimate Mincer regressions of wages on years of schooling and potential work experience, controlling for time and cohort effects of the form:

$$\log w_{ict} = \alpha + \beta s_{ict} + \delta x_{ict} + \gamma_t + \zeta_c + \epsilon_{ict},$$

where w_{ict} is the wage of individual i, who belongs to birth cohort c and is observed at time t. Wages are defined as total annual labor earnings divided by hours worked. We denote by s_{ict} the years of schooling and by x_{ict} work experience, which is defined as

$$x_{ict} = age_{ict} - 18 \text{ if } s_{ict} < 12$$
$$x_{ict} = age_{ict} - s_{ict} - 6 \text{ else.}$$

experience-wage profiles for 4-year experience bins are shown in Table 5, expressed in growth relative to the previous bin. We set the $\{\gamma_{j,E}\}_{j=5}^{16}$ parameters to these values.

Finally, we calibrate the variance of the market luck shocks, σ_{ε}^2 such that our model replicates the standard deviation of labor income across all workers in the data.

Table 5: Human Capital Growth Profiles

Experience (Years)	Wage Growth Non-College	College
0	1.00	1.00
4	0.96	1.15
8	1.09	1.19
12	1.10	1.11
16	1.04	1.06
20	1.02	1.01
24	1.00	0.99
28	1.01	0.97
32	0.99	0.98
36	1.01	0.99
40	0.99	1.01

Notes: This table provides wage growth estimates by year of experience and educational attainment. Source: SOEP

5.3.5 Firms and Government

Following large parts of the literature, we set the capital share in the aggregate production function to $\alpha = 1/3$. Moreover, we set $\sigma_f = 1/3$ such that the elasticity of substitution between vocational and academic human capital in the firm production is equal to 1.5 (Ciccone and Peri, 2005). The weight on vocational human capital in the CES aggregator, ω is estimated internally. Following the arguments in Lee and Seshadri (2019), we calibrate it to match the share of college-educated workers in the SOEP data.

Regarding the tax-related parameters, we set the labor income tax scale to $\lambda = 0.679$ and the labor tax progressivity parameter to $\tau_l = 0.128$ following estimates in Kindermann et al. (2020). The linear capital tax is set to $\tau_a = 0.35$. The size of the lump sum government transfers is set to $g = 0.06$, which in equilibrium amounts to 6% of average labor earnings in the economy. Finally, we set pension benefits to $\pi_j = \Omega h_j w_E$ during retirement and calibrate the scale parameter Ω internally, such that the average replacement rate corresponds to 40%.

To disentangle time from cohort effects, we assume that there is no experience effect on wage growth in the last 8 years of work, following the HLT approach in Lagakos et al. (2018).

Table 6: Parameters calibrated externally

Parameter	Value	Description	Source
Household			
σ	2.0	Inverse EIS	Lee and Seshadri (2019)
γ	0.5	Frisch Elasticity	Fuchs-Schündeln et al. (2022)
q	1.56	Equiv. Scale	Jang and Yum (2022)
$\bar{n}(E=1)$	0.40	Time Cost of College	
Firm			
σ_f	1/3	E.o.S Vocational and Academic Human Capital	Ciccone and Peri (2005)
δ_f	6%	Annual Depreciation	Kindermann et al. (2020)
Government			
τ_n	0.128	Labor Tax Progressivity	Kindermann et al. (2020)
λ	0.679	Labor Tax Scale	Kindermann et al. (2020)
τ_a	0.35	Capital Tax Rate	
g	0.06	Lump-sum Transfers	

Notes: This table presents the externally calibrated parameters and their corresponding sources.

5.4 Method of Simulated Moments Estimation Results

In total, we estimate 20 parameters internally using the method of simulated moments to match 20 target data moments. The parameters, their estimated values, model-implied moments, and target data moments are presented in Table 7.

The model generally fits the data well, both in terms of aggregate moments and concerning the distribution of child skills, school tracks and higher education. For example, the share of of college graduates in the simulated economy is 35.6%, which is in line with the German data in the 2010s. Given that the model also matches the transition rates from academic and vocational secondary school into college higher education (at 70% and 11%), this implies that the share of children in an academic track school in the model, 42% is in accordance with the data.

Parental preferences towards their own track affect the school track decision significantly, both in the model and in the data. In particular, around 20% of parents from each education background overrule a different track recommendation by teachers in the NEPS data. In the model simulated data, roughly the same shares of parents would send their child to a different track if it was not for the own-track bias.

The model is further successful in capturing the transitions between secondary and tertiary education. Around 70% of graduates from an academic track secondary school achieve a college education, while that share is 11% from a vocational track secondary school. Thus,

Table 7: Internally Calibrated Parameters

Parameter	Value	Description	Target	Data	Model
Preferences					
β	0.94	Discount Factor	Annl. Interest Rate	0.04	0.04
b	7	Labor Disutility	Avrg. Labor Supply	0.53	0.53
δ	0.55	Parental Altruism	Transfer/Income	0.60	0.61
χ_0	0.16	Own V-Track Bias	Share of Deviations	0.16	0.19
χ_1	0.17	Own A-Track Bias	Share of Deviations	0.23	0.20
College Costs					
ψ	0.0	Intercept	Share A $\rightarrow$ College	0.71	0.70
ψ_V	0.3	Add. Costs for V-Track	Share V $\rightarrow$ College	0.11	0.11
ψ_θ	0.7	Coefficient on θ_5	Regression Coefficient	0.79	0.8
$\mu_{E^p=0}$	0.18	Mean Taste Shock if $E^p = 0$	Share in CL from Non-CL HH	0.20	0.20
$\mu_{\nu,E^p=1}$	-0.18	Mean Taste Shock if $E^p = 1$	Share in CL from CL HH	0.64	0.64
σ_ν	0.07	Std. Taste Shock	Variance of Residual	0.218	0.163
Idiosyncratic Shocks					
σ_ε	0.0007	Std. Luck Shock	Std(Log Labor Income)	0.73	0.78
σ_ϕ	0.012	Std. Ability Shock	Var(Test Scores Grade 1)	0.022	0.024
ρ_ϕ	0.7	Persistence of Ability	Test Scores Diff. by S	0.46	0.64
σ_{η_3}	0.04	Std. Learning Shock $j = 3$	$\text{Rank}_{j=2}$-$\text{Rank}_{j=3}$	0.57	0.57
σ_{η_4}	0.065	Std. Learning Shock $j = 4$	$\text{Rank}_{j=3}$-$\text{Rank}_{j=4}$	0.61	0.67
σ_{η_5}	0.05	Std. Learning Shock $j = 5$	$\text{Rank}_{j=4}$-$\text{Rank}_{j=5}$	0.67	0.69
Miscellaneous					
Ω	0.145	Pension Anchor	Replacement Rate	0.40	0.40
ξ	5.7	Human Capital Anchor	Avrg. Labor Earnings	1.0	1.0
ω	0.56	Weight V. Human Capital	College Share	0.35	0.36

Notes: This table presents the internally calibrated parameters, targeted moments, and their model-generated counterfactuals.

despite the fact that in principle only an academic school degree qualifies for university entrance, the model can account for the various "second chance" opportunities in the German Education System. In doing so, the college taste shocks play an important role as they ensure that the correlation between parental and child higher education in the model matches the data.

In order to match the correlation between child skill ranks across school periods, the model requires rather large child skill shocks. This is in part, because the estimated own-skill productivity, κ_1, in the child skill formation technology is also quite large. Despite this, the model slightly overstates the differences in initial child skills by parental education prior to entering school. In particular, while children from college-educated parents have an average initial skill level that is around 0.46 standard deviations larger than the average level of non-college-educated parents in the data, this difference is 0.64 standard deviations in the model.

5.5 Validation Exercises

We assess the model's validity using two approaches. First, as is standard in the literature, we compare non-targeted moments from our model simulated data to their counterparts in the NEPS data or using estimates from other research papers. Second, we investigate the effects of school track choice on later-in-life economic outcomes for a set of *marginal* students and compare the results to the null-effects reported in Dustmann et al. (2017) for Germany.

Non-targeted Moments

We summarize selected non-targeted moments and their data or external counterparts in Table 8. The first set of moments pertains to child skills. While we target the difference (in terms of standard deviations) in average initial child skills prior to entering primary school in the calibration, we do not track how this difference evolves over the school career. In the data, the differences in parental education and school track increase slightly during secondary school.[51]

Similarly, the differences in average child skills across school tracks (in terms of standard deviations) increase throughout secondary school, both in the model and in the data. These differences are around twice as large compared to the differences across parental education.

The second set of moments concerns the relationship between track choice and parental education. In the data, the ratio of college-educated parents who choose an academic track

[51] Appendix Table S3 describes the evolution of child skills over time using the identified latent variables.

school for their child relative to the average is 1.46. For non-college-educated parents, this ratio is only 0.81. The model implies a slight overestimation of the first and an underestimation of the second. Moreover, we regress a dummy variable that equals one if a child attends an academic track school on the percentile rank of the child's skills prior to secondary school, in order to assess the skill gradient in academic track choice. The estimated coefficient is 0.75 in the data and 1.42 in the model. Taken together, these moments suggest that our model somewhat overestimates the importance of child skills for the track choice.

The third set of moments relates to intergenerational mobility. To assess the model's validity here, we compare its implications vis-à-vis the estimates on social mobility in Germany reported in Dodin et al. (2021). Using a different data set than we, they regress a dummy of academic-track school graduation of a child on the percentile income rank of her parents, finding that a 10 percentile increase in the parental rank is associated with a 5.2 percentage point increase in the probability of graduating from an academic track school. In our model, a comparable estimate yields a 4.4 percentage point increase. Moreover, Dodin et al. (2021) report absolute graduation rates for children from the first quintile of the income rank distribution (Q1) of 34%, and a ratio of the fifth income rank quintile over the first quintile of 2.13, which our model matches well. We also compare our model-implied estimate of the intergenerational elasticity of income (IGE) to estimates on German data by Kyzyma and Groh-Samberg (2018). Compared to their findings, the model produces IGEs that are at the lower bound of their data counterparts.

Finally, the model understates the degree of inequality in labor incomes as measured by the Gini coefficient. However, the average college wage premium is consistent with the data.

Long-term effects of Track Choice for Marginal Students

Dustmann et al. (2017) analyse the long-term labor market effects of early school track choice in Germany using a quasi-experimental setting. Their identification strategy makes use of the existence of a (fuzzy) cut-off age for school entry in the German system. Children that are born just before the cut-off age are less likely to go to an academic track secondary school, simply because they are younger at the time of the track decision relative to their class peers. This induces a quasi-randomness in secondary school track choice based on the date of birth. The authors then investigate the effect of that date of birth on later-in-life wages, employment and occupation. They find no evidence that the track attended in secondary school affects these outcomes for the marginal children around the school entry cut-off.[52]

[52]Note that Dustmann et al. (2017) control for the effect that being born after the cut-off age directly harms a child's later wages since it means that her labor market entry is later, so that at any given age, she

Table 8: Non-targeted moments

Moment	Data	Model
Child Skill Moments		
Mean Differences by Parental Background (in Standard Deviations)		
Beginning Secondary School	0.48	0.56
Middle Secondary School	0.57	0.47
Mean Differences by School Track (in Standard Deviations)		
Beginning Secondary School	0.87	
Middle Secondary School	1.01	1.10
School Track Choice		
Relative share A-track children from CL. HH	1.46	1.62
Relative share A-track children from Non-CL HH	0.81	0.67
Coefficient A-track on Skill Rank	0.75	1.42
Intergenerational Mobility		
Parental Income Gradient (Dodin et al., 2021)	0.52	0.44
Q5/Q1 A-track on income (Dodin et al., 2021)	2.13	2.30
Q1 A-track on income (Dodin et al., 2021)	0.34	0.28
IGE (Kyzyma and Groh-Samberg, 2018)	0.27-0.368	0.278
Inequality - Returns to College		
Gini Coefficient of Labor Income	0.29	0.22
College Wage Premium	35%	37.6%

Notes: This table presents the non-targeted moments and their model-generated counterfactuals.

We use our model-simulated data to perform a similar exercise. In particular, we are interested in comparing the later-in-life outcomes of children that are very similar in terms of their state variables at the point of school track choice but end up going to different school tracks. Naturally, in our model, we cannot distinguish the date of birth for children of the same cohort. For that reason, we distinguish children by their skills prior to the secondary school track choice ($\theta_{j=3}$). As detailed in Section 4, our child skill development technology implies that, conditional on parental background, the school track choice is characterized by a skill threshold, such that all children with skills above that threshold go to the academic school track and all below go the vocational track school. Conditional on all other states at the time of the track choice – parental human capital, assets, education, and learning ability – differences in child skills and hence differences in school track choice in our model arise from randomly drawn skill shocks. Analogously to Dustmann et al. (2017), we could alternatively argue that these shocks are (at least partly) the result of within-cohort age differences of children, which affect their skill development but are not explicitly modeled. Thus, comparing the later-in-life outcomes of otherwise very similar children with skills

will have accumulated less work experience.

around the tracking threshold can be interpreted as estimating the effect of school track choice induced by random (age or skill) shocks.

Concretely, we compare children with skills in a 5% interval around the tracking threshold who go to different school tracks, conditional on all other states.[53] We evaluate these marginal children in terms of their labor income at age 30, the present value of their lifetime labor income, and the present value of their lifetime wealth.[54] We find that going to the academic track instead of the vocational track is associated with a 6.7% higher labor income at age 30, a 2.2% higher present value of lifetime labor income, and a 4.1% higher present value of lifetime wealth.

While not zero, these differences seem rather small in relation to overall inequality in these outcomes. For example, the 2.2% higher present value of lifetime labor income is around 1/20th of a standard deviation of lifetime labor income. Moreover, in our model, the track choice is only between one vocational and one academic track, whereas Dustmann et al. (2017) consider three tracks, of which two can be classified as vocational. We would generally expect that children at the margin of these two vocational tracks show fewer differences in lifetime outcomes. In sum, we conclude that the implications our model entails with respect to the effect of tracking on *marginal* children are not at odds with the reduced-form evidence presented in Dustmann et al. (2017).

6 Quantitative Results

The benefit of our model is that we can use it to understand the effects of school tracking not only for marginal children but for the whole distribution of children, their educational and labor market outcomes, as well as their economic mobility relative to their parents. To that end, we first use our model to quantify the sources of lifetime and inter-generational inequality in the spirit of Huggett et al. (2011) and Lee and Seshadri (2019). Then, we investigate the determinants and consequences of secondary school track choice, as this constitutes the main novelty of our model. In this context, we perform counterfactual analysis of economies in which the school track decision is not affected by an own-track bias of the

[53] This interval amounts to around 1/5 of a standard deviation of child skills prior to the school track choice. We form quintiles of the continuous states of parental human capital and parental assets and allocate children into discrete groups pertaining to these quintiles. Moreover, we partition the distribution of the learning ability ϕ' into three ability states. For these reasons, the skill threshold can become fuzzy in the sense that even conditional on these groups a child with slightly higher skills goes to the vocational track whereas a child with slightly lower skills goes to the academic track.

[54] Lifetime labor income is computed as the discounted sum of all labor income during the adult periods, and lifetime wealth is that sum plus the initial monetary transfer from the parent to their independent child.

parents or in which a policymaker enforces a strict tracking skill threshold. Finally, we study the effects of a counterfactual policy reform that postpones the school tracking age to 14.

6.1 Sources of Inequality

Using our model, we can decompose how much of the variation in lifetime economic outcomes of our model agents can be explained by various factors at various ages. Following the literature, we focus on lifetime labor income and lifetime wealth as our economic outcomes of interest. We begin by computing the contribution of each state variable of a freshly independent child at age 18 to the variation in lifetime labor income and wealth.[55] These states are the school track in secondary school, S, initial adult human capital h_5, initial transfers received from the parent a_5, parental education E^p, and innate learning ability ϕ.

Table 9: Contributions to Lifetime Inequality

		Share of Explained Variance	
Life Stage	States	Lifetime Earnings	Lifetime Wealth
Independence (age 18)	(S, ϕ, h_5, a_5, E^p)	80%	74%
	(S, ϕ, h_5, E^p)	75%	72%
	(S, ϕ, a_5, E^p)	57%	44%
School Track Choice (age 10)	$(S', \phi', \theta'_3, h_{11}, a_{11}, E)$	21%	25%
	(S', θ'_3, ϕ')	20%	22%
	(S')	15%	19%
Pre-Birth (parent age 30)	(E, ϕ, h_8, a_8)	4.5%	11.3%

Notes: This table shows how much of the variation in lifetime economic outcomes is explained by different factors at different ages.

Row 1 of Table 9 summarizes that 80% of the variation in lifetime labor income can be accounted for by all states at the age of 18. In terms of lifetime wealth, this number is around 74%.[56] Thus, our model suggests that the majority of lifetime outcomes is already predetermined when agents become independent and can enter the labor market. Note that at this stage, all uncertainty regarding initial human capital as well as the college decision has been made. The remaining unresolved uncertainty over human capital (market luck) shocks during the working years has, therefore, only limited effects on lifetime inequality.

[55] Concretely, we follow the approach in Lee and Seshadri (2019) and calculate conditional variances of lifetime labor income and wealth, after conditioning on the state variables. As before, we partition the continuous states into three equally sized groups.

[56] These numbers are comparable with estimates for the U.S. (Lee and Seshadri, 2019; Huggett et al., 2011; Keane and Wolpin, 1997)

As Row 2 of Table 9 shows, the explained share of variation in lifetime outcomes remains high if we only condition on the states that directly pertain to the newly-independent child. In particular, the size of the parental transfer a_5 and parental education, which affects the college taste shock, do not independently contribute much to lifetime inequality. This changes if we only exclude initial adult human capital h_5. The share of explained variance in lifetime earnings drops by almost 20 percentage points, and the share of explained variance in lifetime wealth by almost 30 percentage points. This suggests that variation in human capital, even at age 18, is an important driver of lifetime inequality.[57] Interestingly, the correlation between initial adult human capital and transfers received from parents is negative in the model. This suggests that parents partially offset the disadvantage their children experience in the labor market from having lower skills by giving them higher transfers.[58]

Using the same methodology, we can also evaluate how much lifetime inequality is already determined at the time of the school track choice. Conditioning on all states at that age, around 21% of lifetime earnings and 25% of lifetime wealth variation is explained. Again, the majority of this variation seems attributable to differences in child states at that age. Yet the explained share is clearly smaller than after school, suggesting that the learning outcomes during the secondary school play an important role in shaping later-in-life inequality. Conditioning on the initial school track choice alone can account for 15% of lifetime earnings variation and 19% of lifetime wealth variation. However, this should not be interpreted as the marginal effect of school track choice on lifetime outcomes, as the initial school track choice is, for example, highly correlated with child skills at that age. In fact, as we argued in Section 5.5, for children with similar skills, the track choice has only small independent effects on lifetime outcomes. We investigate the determinants and consequences of the school track choice in more detail below.

The last row of Table 9 shows the contribution of parental states prior to the birth of their children to their children's lifetime outcomes. At this stage, none of the uncertainty regarding child skill and human capital shocks nor regarding the child's learning ability has been realized. Still, around 4.5% of the variance in lifetime earnings of the yet-to-be-born child is predetermined by parental education, ability, human capital, and wealth. For lifetime wealth, this share is even higher at 11.3%, pointing to the important role of wealth transfers. For example, suppose we use the same decomposition of the unconditional variance of transfers into parental states pre-birth. In that case, we find that almost 39% of variation

[57]We cannot, however, attribute these drops exclusively to human capital differences, given the possible correlation between states.

[58]This channel is also present, albeit to a smaller degree, in Lee and Seshadri (2019).

in transfers is predetermined prior to the birth of the child. In contrast, only 6.2% of the variation in human capital at age 18 is predetermined prior to birth, which highlights the role of shocks to child skills during their childhood and school years.

6.2 School Track Choice Counterfactuals

According to the theoretical predictions laid out in Section 4, the initial school track should, to a large degree, be based on child skills. A regression of an academic school track dummy on all states at the time of the tracking decision confirms that this is true. Column 1 of Table 10 reports the standardized coefficient estimates of this regression, indicating that child skills at the time of the track choice, θ_3' have the strongest impact on the track decision. In particular, increasing log child skills by one standard deviation increases the probability of going to the academic track by 70 percentage points.

Notwithstanding this, Column 1 in Table 10 also indicates that parental education is the second most important independent driver of the school track choice. In the model, parental education can influence the track choice, net of the effects coming through child skills, human capital, or wealth, in three ways. First, college-educated parents know that their children learn faster than their non-college-educated counterparts. This comes from the estimated direct parental education effect in the child skill production technology, κ_5. This knowledge may prompt college parents to send their child to the academic track even if their child's skills are lower than those of a child from a vocational parent. Second, parents know that their child will receive a college taste shock that depends on their parent's education, governed by μ_{ν,E^p}. In anticipation of this, college parents, for instance, may have a stronger incentive to send their child to an academic track school as this, everything else equal, increases the likelihood of college admission. However, (non-pecuniary) college costs also decrease in end-of-school skills. As derived in Section 4, for a set of children with low preschool skills, end-of-school skills are maximized if they attend the vocational school track. This force counteracts the incentive of college parents to send their child to the academic track described before. Third, even net of college tastes, we assume that parents bias the school track choice towards their own education level. We motivated this bias by the significant number of deviations from teacher recommendations in the school track choice. The bias then directly implies a stronger direct effect of parental education on the school track choice.

To understand how important each of these channels for the school track choice is, we perform a series of three counterfactual experiments using the calibrated model, in which

Table 10: School Track Choice Determinants

	Dependent Variable: $S = A$			
	Stand. Coefficient Estimates			
	(1)	(2)	(3)	(4)
	Baseline	$\kappa_{5,j=3,4} = 0$	$\mu_{\nu,1} = \mu_{\nu,0} = 0$	$\chi_0 = \chi_1 = 0$
ϕ'	0.033	0.025	0.021	0.024
θ_3'	0.708	0.710	0.717	0.768
$E = 1$	0.174	0.159	0.163	0.037
h_{11}	0.032	0.034	0.034	0.024
a_{11}	0.009	0.007	0.010	0.007

Notes: This table reports the standardized coefficient estimates of regressions of an academic school track dummy on all states at the time of the tracking decision. Column (1) corresponds to the baseline economy. In Column (2), we shut down the channel of differential parental inputs in periods 3 and 4. Column (3) considers the case of identical college taste shock by parental education. In Column (4), we remove the parental preference bias for education.

we isolate each effect, respectively.[59] In particular, we isolate the effects of the first channel by solving the model with $\kappa_{5,j=3,4} = 0$ yet leaving $\kappa_{5,j=3,4} > 0$ in the simulation of the distribution. That is, we assume that parents do not take into account the direct effect of their own education on child skill development during secondary school when making the track decision. The skills, however, still evolve as in the baseline model. Column (2) in Table 10 reports the (standardized) results of the regression of academic track choice on all state variables in this counterfactual scenario. The coefficient on parental education drops as expected, while the coefficient on child skills prior to the track decision increases. This confirms that the knowledge of direct parental effects on future child skill development prompts parents to send their child to the same track as their own, net of effects of parental education through child skills that are already formed. The magnitude of this channel, however, seems relatively small. In particular, the results suggest that this channel accounts for around 8.5% of the direct effect of parental education on the probability of academic school track attendance of her child.

Column (3) reports the resulting coefficient estimates when isolating the second channel, working through college tastes. If we equalize the means in college taste shocks across parental education (to zero), once again, the coefficient on direct parental influence on school track choice decreases, and the one on child skill increases. However, this effect is quantitatively even less important than the first channel and can account for around 6% of the

[59] In doing so, we again solve for the stationary general equilibrium allowing prices to clear the markets and average child skills across tracks to be consistent with the parents' track decision.

parental influence. This may likely be because of the complex college cost structure, which depends both on the school track, but also on the end-of-school skill level, as explained above.

The most important reason for the remaining influence of parental background on a child's school track works through the immediate own-track bias governed by χ_E. As reported in Column (4) of Table 10, the direct influence of parental education on the school track of a child drops by almost 80% if we set $\chi_E = 0$ for both education levels. Similarly, the effect of parental human capital at the time of the track choice drops. At the same time, a child's own skills become more important for the track decision. As discussed at the end of Section 4, this suggests that preferences for own tracks by parents can create inefficiencies in the allocation of children across school tracks.[60]

An important question is whether the consequences of such misallocation effects are visible not only in terms of child skill outcomes but also in the aggregate and distributional outcomes in the economy. Our model provides a suitable environment to investigate such effects. Table 11 provides an overview of selected outcomes in the baseline model (Column (1)) and compares the resulting percentage change of these outcomes in the counterfactual scenario without own-track biases in the school track choice (Column (2)). Moreover, we report in Column (3) the relative changes in the outcomes in another counterfactual experiment. We enforce that the school track choice is governed exclusively by a sharp skill threshold. This threshold is chosen such that the bottom 50% of children are allocated to the vocational track, while the top 50% go to the academic track, regardless of their parental background. As derived in Section 4, this constitutes the optimal tracking policy from the point of view of a policymaker who is only interested in maximizing aggregate end-of-school skills and cannot condition on the parental background.[61]

In both counterfactual scenarios, the share of college-educated agents in the economy drops slightly relative to the baseline case. Moreover, aggregate output is virtually the same in all three economies. This is perhaps surprising as the share of children that attend an academic track school increases. In the case without preference-based school track choice,

[60]Suppose college-educated parents send their children to an academic track school, despite the fact that their skill level would optimally suggest the vocational track. In that case, this will not only harm their child's development but also cause the instruction pace in that track to adjust. This, in turn, harms the average learning gains of everyone in that track. The same effect occurs in the vocational track school if parents from non-college backgrounds send their overqualified children there purely based on preferences.

[61]The optimal tracking threshold derived in Section 4 is exactly at the average child skill level prior to the track decision. Picking this threshold ensures that the variance of child skills in each track is minimized. When child skills are normally distributed, this results in a 50:50 split of children across tracks. In our model economy, the log of child skills is approximately normally distributed, with a mean slightly lower than 0.

the share increases by 4%. By construction, this share increases even further in the case of the sharp track threshold, as this threshold is picked such that 50% of the children go to either track. The reason for the non-existent effects on output and college education becomes clearer when we study the distribution of skills in counterfactual experiments.

In particular, the first two rows in Panel B. of Table 11 suggest that both counterfactual scenarios lead to an increase in average child skills in the middle and end of secondary school. This increase arises from the fact that in both school tracks, the variation in child skills is smaller than in the baseline economy, as can be seen in the last four rows of Panel B. Thus, peer groups are more homogeneous if we take away parental preferences in the track choice and even more so if we employ a strict skill threshold rule for the track allocation. This is consistent with the explanation of the efficiency-reducing misallocation effects that arise when parental background affects the school track choice, as explained above.

However, the magnitude of the increases in efficiency is very small at 0.2% in period $j = 4$ and becomes even smaller at the end of school with 0.1%. The reason for this small effect is likely a combination of the quantitatively relatively small role of parental background, *net of* child skills, in driving the school track choice even in the baseline economy (see Table 10) and the relatively large child skill shock sizes that undermine the importance of homogeneous peer groups for learning efficiency. This also explains why the average skill level gains even decrease over time. Given that children in both counterfactual scenarios are, on average, only slightly more skilled than in the baseline economy. Prices adjust such that the labor supply of both college and non-college labor meets the demand by the representative firm. The incentives to go to college also seem comparable to the baseline economy, despite the larger share of academic track children.

Rows 3-7 of Panel A. suggest that without parental biases in school track choice and even more so with a sharp, purely skill-based allocation rule, the dependence of school track choice on parental income decreases. Moreover, skills themselves become more important in explaining the track choice and the college choice. Unsurprisingly, the intergenerational elasticity between parents' and child's income drops in both cases.

Overall, Table 11 paints the following picture. Both counterfactual scenarios achieve an improvement in child learning during the secondary school years. However, this improvement is not sufficient to yield a significant effect in terms of aggregate output in the macroeconomy. At the same time, both counterfactual experiments result in more upward mobility of children relative to their parents.

Table 11: Effects of School Track Choice Counterfactuals

Outcome	(1) Baseline Economy	(2) $\chi_0 = 0$ $\chi_1 = 0$	(3) Sharp Threshold s.t. 50:50 Split
		Panel A.	
Y	1.11	0.0%	0.0%
College Share	0.36	-0.8%	-0.6%
A-Track Share	0.42	+4.0%	+18.8%
A-Track on Income	0.44	-33.1%	-45.3%
A-Track on Skills	1.42	+4.0%	+5.8%
CL on Skills	0.79	+0.3%	+1.9%
IGE	0.28	-3.6%	-1.8%
Gini Earnings	0.22	+0.5%	0.0%
		Panel B.	
$\bar{\theta}'_4$	0.81	+0.2%	+0.2%
$\bar{\theta}'_5$	0.79	+0.1%	+0.1%
$Std(\theta'_{4\|S'=V})$	0.21	-1.9%	-5.2%
$Std(\theta'_{4\|S'=A})$	0.33	-0.9%	-2.4%
$Std(\theta'_{5\|S'=V})$	0.28	-0.4%	-1.1%
$Std(\theta'_{5\|S'=A})$	0.28	-0.4%	-1.4%

Notes: Column (1) shows aggregate outcomes in the baseline economy, Column (2) displays percentage changes entailed by the absence of parental preference for education, and Column (3) displays percentage changes entailed by a 50-50 rule for school tracking.

6.3 Postponing the School Tracking Age

An important feature of school tracking policies is the age at which children are allocated across the tracks. Generally, OECD countries differ remarkably in the school tracking age (see Figure IV.2.2 in OECD (2013) for an overview). In countries with an early tracking system in place, such as Germany, it is often argued that postponing the tracking age will improve equality of opportunity in terms of access to academic education without incurring efficiency losses in terms of learning outcomes (Woessmann, 2013). While some reduced-form estimates, exploiting cross-country, federal-state level, or time differences in tracking policies exist, little is known about the aggregate, distributional, and inter-generational consequences of a large-scale reform that postpones the tracking age.

To evaluate such a reform in the context of Germany, we conduct a series of Late Tracking counterfactual experiments using our calibrated model. In each experiment, we assume that the age at which children can sort into an academic or vocational school track is postponed from 10 to 14, corresponding to model period $j = 3$ to $j = 4$. During model period $j = 3$, all children attend a school that belongs to a comprehensive school track, just like during primary school in $j = 2$. In each counterfactual experiment, all parameters, in particular those governing school track preferences and college costs, remain the same as in the baseline economy.

We present the relative changes of selected aggregate and social mobility outcomes of the counterfactual experiments relative to the baseline economy in Table 12. The experiments differ in the way we assume that prices and instruction paces are allowed to adjust. In Column (2), all prices (wages per efficiency unit for college and non-college human capital w_0, w_1 and the interest rate r) are assumed to remain at the same values as in the baseline case. That is, we compare the partial equilibrium outcomes of the policy counterfactual. Moreover, we assume that the instruction pace during the *second stage* of secondary school does not adjust. That is, the policymaker sets the same pace as in the baseline case in both academic and vocational track schools during $j = 4$. As a result, parents do not need to form expectations over the average skill levels in each track when they make the postponed track choice.

In this economy, aggregate output Y is around 3.6% lower than in the baseline case. This is despite the fact that the share of college-educated agents increases by almost 50% and the share of children in the academic track in $j = 4$ even increases by 75%. However, average human capital is significantly less than in the baseline economy, which is ultimately a result of less efficient learning during secondary school. In particular, average child skills in period

$j = 4$ are almost 8% lower in the late tracking case than in the baseline economy. At the end of school, skills are still over 4% lower than in the baseline case. Thus, even though wages per efficiency unit is unchanged, the college premium that measures earnings divided by labor supply declines.

As we derived in Section 4, it is theoretically not clear whether later tracking results in such learning efficiency losses. In particular, later tracking could even increase average learning outcomes if the variance of the child skill shocks is sufficiently large. The reason for that is that with large skills shocks, the gain from more homogeneous peer groups in each track during the last stage of secondary school can out-weight the losses incurred due to one more period of learning in a comprehensive track during the first stage of secondary school. These effects are also visible in Panel B. of Table 12. The standard deviation of child skills in both tracks in model period $j = 4$ is significantly smaller compared to the baseline case. However, this smaller variation within school tracks cannot overcome the disadvantage in terms of average skills with which children enter into period $j = 4$. Thus, despite sizable estimates of the child skill shocks variances, our model predicts that the learning losses from postponing 4 years of tracking in Germany cannot be recuperated by more-efficient learning during the remainder of secondary school.

It could be, of course, that later tracking aids in resolving another type of inefficiency that is present in the baseline early tracking economy: that of misallocation due to biased-track choices by parents. This would be the case if going to the academic track depended less on the income or education of the parent after the late tracking policy reform. While this is generally true, as can be seen in Table 12, it is also true that academic track attendance depends even less on the child's skill level prior to the track choice than in the early tracking case. Fundamentally, this is just a result of the larger share of children that go to the academic track in the first place. As a consequence, it can be said that inequality of opportunity in access to academic secondary education is reduced as the academic track simply becomes the preferred track for the majority of parents, regardless of their background.

This pattern becomes even more striking if we allow the instruction pace in each track to adjust endogenously in Column (3) while still keeping prices at their baseline values. In this case, almost every child ends up in the academic track school, which harms aggregate learning even more. Even though around 70% of agents in the economy then end up having a college degree, aggregate output is still almost 10% lower than in the early tracking case. The high share of college-educated agents is, of course, dependent on keeping the effective wages fixed at their early tracking value. Once we relax this assumption in Columns (4) and (5), the college share increases by only 2.2% and 4.2%, respectively. Yet, despite the fact that the distribution of higher education across parents is similar to the early tracking baseline, the share of children that go to the academic track still increases substantially. Indeed, if we allow for an adjusting instruction pace, still more than 90% of children go to the academic track school.

This signals that, in the late tracking counterfactual, the economic forces that link the secondary school track decision to the eventual college decision are much weaker than in the early tracking case. A reason for this could be that, in terms of the model, a child skill shock realization during $j = 4$ that renders a school track choice ex-post inefficient (for example a late-bloomer shock), can be corrected already in the next period when it is possible to switch tracks again. Thus, when parents make the school track choice, they put more relative focus on purely maximizing end-of-school skills, regardless of the track. Another part of the explanation is the fact the parameter estimates of the child skill formation technology in the second stage of secondary school (see the estimates for $j = 4$ in Table 4) put less emphasis on the distance to the average peer level compared to $j = 3$. This reduces the learning penalty of sending your child to the academic track.

Both effects are amplified when the teaching pace adjusts endogenously, as fewer and fewer parents find the vocational track preferable which lowers the instruction pace even further. The resulting equilibrium finally resembles a schooling system that is almost completely comprehensive throughout all years but features a special track for a small share of especially low-skilled children at age 14. Naturally, a side effect of such a system is that skills at the end of school become more equal. This translates not only into smaller overall inequality in the economy, as exemplified by a 3.6% smaller Gini coefficient of labor income, but also into significantly higher upward mobility. In particular, the inter-generational income elasticity is reduced by over 11% in the general equilibrium case with adjusting instruction pace.

The main takeaways of the policy reform that postpones school tracking to age 14 in our model can be summarized as follows. First, postponing school tracking incurs efficiency losses from worse learning outcomes in the additional period of comprehensive school. The losses cannot be compensated by gains in later years that arise from more homogeneous peer groups across tracks as the track decision is based on more complete information about children's true potential. Second, later tracking incentivizes more parents to send their child to an academic track secondary school, in particular, if the instruction pace is set to reflect the distribution of children across tracks that arises from the track allocation game

Table 12: Late Tracking Counterfactuals

Outcome	(1) Baseline	(2) Baseline pace + PE	(3) Pace adjusts + PE	(4) Baseline pace + GE	(5) Pace adjusts + GE
			Panel A.		
Y	1.11	-3.6%	-9.8%	-2.1%	-3.3%
College Share	0.36	+48.3%	+73.6%	+2.2%	+4.2%
A-Track Share	0.42	+74.6%	+131.4%	+63.9%	+127.6%
College Premium	1.38	-4.0%	-3.8%	-4.8%	-7.5%
Gini Earnings	0.22	0.0%	+0.5%	-2.3%	-3.6%
A-Track on Income	0.44	-7.2%	-86.9%	-16.7%	-86.7%
A-Track on Skills	1.42	-44.6%	-90.6%	-40.2%	-85.9%
CL on Skills	0.79	+1.9%	-12.2%	-2.5%	-7.3%
IGE	0.28	-8.6%	-14.7%	-8.3%	-11.5%
			Panel B.		
$\bar{\theta}'_4$	-0.22	-7.9%	-3.3%	-12.6%	-12.6%
$\bar{\theta}'_5$	-0.23	-4.3%	-7.3%	-8.2%	-12.4%
$Std(\theta'_{4\|S'=V})$	0.21	-11.4%	-41.0%	-10.5%	-37.6%
$Std(\theta'_{4\|S'=A})$	0.33	-12.8%	-9.8%	-14.7%	-12.8%
$Std(\theta'_{5\|S'=V})$	0.28	-8.8%	-13.4%	-8.5%	-31.8%
$Std(\theta'_{5\|S'=A})$	0.28	-5.0%	-7.2%	-6.8%	-9.3%

Notes: This Table presents changes in outcomes due to delaying the school tracking choice by four years (from the age of ten to the age of fourteen). Column (1) shows aggregate outcomes in the baseline economy, and Columns (2) to (5) display percentage changes due to the policy change in different scenarios. Column (2): if the pace of instruction and prices are unchanged. Column (3): if the pace of instruction adjusts but prices are unchanged. Column (4): if the pace of instruction is unchanged but prices adjust. Column (5): if the pace of instruction and prices adjust.

of the parents. Moreover, the track decision seems largely separate from college admission concerns.[62] Third, this results in more equal access to academic secondary education, which leads to more equal access to higher education and more equal labor market outcomes. It also reduces the persistence of economic status across generations.

7 Conclusion

How important is the design of education policies for the macroeconomic analysis of inequality and social mobility? This paper argues that school tracking, a standard policy across many advanced countries, influences not only equality of educational opportunities for children from different parental backgrounds but also shapes aggregate learning and, as a consequence, aggregate economic efficiency. We add a macroeconomic perspective to the predominantly reduced-form literature by building a macroeconomic GE model of overlapping generations that specifically zooms in on the children's schooling years. To that end, we formulate a simple theory of child skill formation, where child skills depend linearly on her classroom peers and non-linearly on the instruction pace specific to each school track.

We show that this child skill formation technology alone entails theoretical implications for the effect of school tracking policies on the distribution of child skills that align with the most robust findings of a vast empirical literature and the most popular arguments in the public debate about tracking. In particular, not every child gains from tracking; the losses are often concentrated among lower-skilled children. Additionally, tracking can lead to increased inequality in end-of-school skills. Finally, the effects of tracking on learning efficiency, while typically positive on average, depend on the age at which children are tracked and the size of uncertainty regarding the evolution of child skills, highlighting the importance of the timing of tracking.

We embed this theory into a standard Aiyagari-style life-cycle framework in which parents make a school track decision for their children. We tailor the model to fit the German Education System, where the track decision occurs at the age of 10 of the child, and calibrate it on German data. Our quantitative results suggest that variation from the initial school track alone can account for around 15% of the variation in eventual lifetime earnings. Conditional on prior child skills, the track choice is strongly influenced by parental preferences for their track that parental inputs into child skills or tastes for higher education cannot explain. This gives rise to efficiency-reducing misallocation of children across tracks. However,

[62] This result seems counterfactual. Further work on modeling assumptions is needed.

we find that the magnitudes of this efficiency loss are small and completely wash out in the later life stages. Thus, even a tracking policy that splits children solely based on their skills while improving social mobility does not lead to meaningful aggregate output gains in the macroeconomy. Our paper also shows that a policy reform that delays the school tracking decision by four years (to age 14) entails aggregate output losses in the long run that amount to around 2-3% of GDP while decreasing the inter-generational income elasticity by 8-11%. This is because such a reform results in many children being allocated into one track, which makes everyone more equal but harms learning efficiency.

References

Aakvik, A., Salvanes, K. G., and Vaage, K. (2010). Measuring heterogeneity in the returns to education using an education reform. *European Economic Review*, 54(4):483–500.

Abbott, B., Gallipoli, G., Meghir, C., and Violante, G. L. (2019). Education policy and intergenerational transfers in equilibrium. *Journal of Political Economy*, 127(6):2569–2624.

Agostinelli, F. (2018). Investing in children's skills: An equilibrium analysis of social interactions and parental investments. *Unpublished Manuscript, University of Pennsylvania.*

Agostinelli, F., Doepke, M., Sorrenti, G., and Zilibotti, F. (2023). It takes a village: the economics of parenting with neighborhood and peer effects. Working Paper w27050, National Bureau of Economic Research.

Agostinelli, F., Saharkhiz, M., and Wiswall, M. (2019). Home and school in the development of children. Working Paper w26037, National Bureau of Economic Research.

Agostinelli, F. and Wiswall, M. (2016). Estimating the technology of children's skill formation. Working Paper w22442, National Bureau of Economic Research.

Ballou, D. (2009). Test scaling and value-added measurement. *Education finance and Policy*, 4(4):351–383.

Becker, G. S. and Tomes, N. (1986). Human capital and the rise and fall of families. *Journal of labor economics*, 4(3, Part 2):S1–S39.

Bellenberg, G. and Forell, M. (2012). Schulformwechsel in deutschland. *Durchlässigkeit und Selektion in den*, 16.

Blossfeld, H., Roßbach, H., and von Maurice, J. (2011). The german national educational panel study (neps). *Zeitschrift für Erziehungswissenschaft: Sonderheft*, 14.

Bonesrønning, H., Finseraas, H., Hardoy, I., Iversen, J. M. V., Nyhus, O. H., Opheim, V., Salvanes, K. V., Sandsør, A. M. J., and Schøne, P. (2022). Small-group instruction to improve student performance in mathematics in early grades: Results from a randomized field experiment. *Journal of Public Economics*, 216:104765.

Borghans, L., Duckworth, A. L., Heckman, J. J., and Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of human Resources*, 43(4):972–1059.

Capelle, D. (2022). The great gatsby goes to college: Tuition, inequality and intergenerational mobility in the u.s. Working paper.

Caucutt, E. M. and Lochner, L. (2020). Early and late human capital investments, borrowing constraints, and the family. *Journal of Political Economy*, 128(3):1065–1147.

Ciccone, A. and Peri, G. (2005). Long-run substitutability between more and less educated workers: evidence from us states, 1950–1990. *Review of Economics and statistics*, 87(4):652–663.

Cunha, F. and Heckman, J. (2007). The technology of skill formation. *American economic review*, 97(2):31–47.

Cunha, F., Heckman, J., and Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931.

Daruich, D. (2022). The macroeconomic consequences of early childhood development policies. Working Paper 2018-29, FRB St. Louis.

Dodin, M., Findeisen, S., Henkel, L., Sachs, D., and Schüle, P. (2021). Social mobility in germany. Discussion Paper DP16355, CEPR.

Duflo, E., Dupas, P., and Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *American Economic Review*, 101(5):1739–74.

Dustmann, C. (2004). Parental background, secondary school track choice, and wages. *Oxford Economic Papers*, 56(2):209–230.

Dustmann, C., Puhani, P. A., and Schönberg, U. (2017). The long-term effects of early track choice. *The Economic Journal*, 127(603):1348–1380.

Epple, D. and Romano, R. (2011). Peer effects in education: A survey of the theory and evidence. In *Handbook of social economics*, volume 1, pages 1053–1163. Elsevier.

Falk, A., Kosse, F., and Pinger, P. (2021). Mentoring and schooling decisions: Causal evidence. Working Paper 8382, CESifo.

Fuchs-Schündeln, N., Krueger, D., Kurmann, A., Lale, E., Ludwig, A., and Popova, I. (2023). The fiscal and welfare effects of policy responses to the covid-19 school closures. *IMF Economic Review*, pages 1–64.

Fuchs-Schündeln, N., Krueger, D., Ludwig, A., and Popova, I. (2022). The long-term distributional and welfare effects of covid-19 school closures. *The Economic Journal*, 132(645):1647–1683.

Fujimoto, J., Lagakos, D., and Vanvuren, M. (2023). Aggregate and distributional effects of 'free' secondary schooling in the developing world. Working Paper w31029, National Bureau of Economic Research.

Guyon, N., Maurin, E., and McNally, S. (2012). The effect of tracking students by ability into different schools a natural experiment. *Journal of Human resources*, 47(3):684–721.

Hanushek, E. A. and Wössmann, L. (2006). Does educational tracking affect performance and inequality? differences-in-differences evidence across countries. *The Economic Journal*, 116(510):C63–C76.

Heathcote, J., Storesletten, K., and Violante, G. L. (2017). Optimal tax progressivity: An analytical framework. *The Quarterly Journal of Economics*, 132(4):1693–1754.

Heckman, J. and Mosso, S. (2014). The economics of human development and social mobility. *Annu. Rev. Econ.*, 6(1):689–733.

Henninges, M., Traini, C., and Kleinert, C. (2019). Tracking and Sorting in the German Educational System. Working Paper 83, Leibniz Institute for Educational Trajectories (LIfBi).

Huggett, M., Ventura, G., and Yaron, A. (2011). Sources of lifetime inequality. *American Economic Review*, 101(7):2923–54.

Jang, Y. and Yum, M. (2022). Aggregate and Intergenerational Implications of School Closures: A Quantitative Assessment. Working Paper 234v1, CRC TR 224.

Keane, M. P. and Wolpin, K. I. (1997). The career decisions of young men. *Journal of political Economy*, 105(3):473–522.

Kindermann, F., Mayr, L., and Sachs, D. (2020). Inheritance taxation and wealth effects on the labor supply of heirs. *Journal of Public Economics*, 191:104127.

Kyzyma, I. and Groh-Samberg, O. (2018). "intergenerational economic mobility in germany: Levels und trends ". Working paper, DIW.

Lagakos, D., Moll, B., Porzio, T., Qian, N., and Schoellman, T. (2018). Life cycle wage growth across countries. *Journal of Political Economy*, 126(2):797–849.

Lee, S. Y. and Seshadri, A. (2019). On the intergenerational transmission of economic status. *Journal of Political Economy*, 127(2):855–921.

Malamud, O. and Pop-Eleches, C. (2011). School tracking and access to higher education among disadvantaged groups. *Journal of Public Economics*, 95(11-12):1538–1549.

Matthewes, S. H. (2021). Better together? heterogeneous effects of tracking on student achievement. *The Economic Journal*, 131(635):1269–1307.

Meghir, C. and Palme, M. (2005). Educational reform, ability, and family background. *American Economic Review*, 95(1):414–424.

Nennstiel, R. (2022). No matthew effects and stable ses gaps in math and language achievement growth throughout schooling: Evidence from germany. *European sociological review*.

Neumann, I., Duchhardt, C., Grüßing, M., Heinze, A., Knopp, E., and Ehmke, T. (2013). Modeling and assessing mathematical competence over the lifespan. *Journal for educational research online*, 5(2):80–109.

OECD (2013). Pisa 2012 results: What makes schools successful (volume iv): Resources, policies and practices, pisa. Report TD/TNC 114.1481, OECD.

Passaretta, G., Skopek, J., and van Huizen, T. (2022). Is social inequality in school-age achievement generated before or during schooling? a european perspective. *European Sociological Review*, 38(6):849–865.

Pekkala Kerr, S., Pekkarinen, T., and Uusitalo, R. (2013). School tracking and development of cognitive skills. *Journal of Labor Economics*, 31(3):577–602.

Piopiunik, M. (2014). The effects of early tracking on student performance: Evidence from a school reform in bavaria. *Economics of Education Review*, 42:12–33.

Pohl, S. and Carstensen, C. H. (2012). Neps technical report-scaling the data of the competence tests. Working Paper 14, NEPS.

Pohl, S. and Carstensen, C. H. (2013). Scaling of competence tests in the national educational panel study-many questions, some answers, and further challenges. *Journal for Educational Research Online*, 5(2):189–216.

Ruhose, J. and Schwerdt, G. (2016). Does early educational tracking increase migrant-native achievement gaps? differences-in-differences evidence across countries. *Economics of Education Review*, 52:134–154.

Schneider, T. and Linberg, T. (2022). Development of socio-economic gaps in children's language skills in germany. *Longitudinal and Life Course Studies*, 13(1):87–120.

Woessmann, L. (2013). Die entscheidende säule. *Wirtschaftswoche Global*, 2(24.06.2013):110–111.

Woessmann, L. (2016). The importance of school systems: Evidence from international differences in student achievement. *Journal of Economic Perspectives*, 30(3):3–32.

Yum, M. (2022). Parental time investment and intergenerational mobility. *International Economic Review*.

A Appendix

A.1 Proof of Propositions

In the following, $\delta = \gamma$.

Proposition 1

First, we show that maximizing the aggregate end-of-school skills in a tracking system implies a threshold skill level $\tilde{\theta}_1$, such that all $\theta_1 < \tilde{\theta}_1$ go to one track, call it $S = V$ and all $\theta_1 > \tilde{\theta}_1$ go to the other track, $S = A$ (and those with $\theta_1 = \tilde{\theta}_1$ are indifferent). That is, the existence of a skill threshold is a necessary condition for optimal end-of-school skills. We restrict ourselves to the case with different instruction paces across school tracks.

To that end, it is useful to rewrite θ_2 in (22) of a child in a given school track S with instruction pace P_S^* using Lemma 1 as:

$$\theta_2 = \theta + \alpha\bar{\theta}_S + \frac{\beta^2}{2\delta} + \frac{\beta\gamma\theta_1}{\delta} + \frac{\gamma^2\theta_1\bar{\theta}_{S_c}}{\delta} - \frac{\gamma^2\bar{\theta}_{S_c}^2}{2\delta} + \eta_2. \tag{26}$$

After adding and subtracting $\frac{\gamma^2}{2\delta}\theta_1^2$, this can be expressed as

$$\begin{aligned}\theta_2 &= \theta_1 + \alpha\bar{\theta}_S + \frac{\beta^2}{2\delta} + \frac{\beta\gamma\theta_1}{\delta} + \frac{\gamma^2\theta_1^2}{2\delta} + \eta_2 - \frac{\gamma^2}{2\delta}\left(\theta_1^2 - 2\theta_1\bar{\theta}_{S_c} + \bar{\theta}_{S_c}^2\right) \\ &= \theta_2(P^*(\theta_1)) - \frac{\gamma^2}{2\delta}(\theta_1 - \bar{\theta}_S)^2,\end{aligned} \tag{27}$$

where $\theta_2(P^*(\theta_1))$ denotes end-of-school skills if the child with skills θ_1 is taught at her individually optimal teaching pace $P^*(\theta_1)$. Thus, in a given track end-of-school skills are a strictly decreasing function of the *distance* to the average skill level $\bar{\theta}_S$ in that track. This is intuitive given Lemma 1, as it is solely the average skill level to which the instruction pace is optimally targeted.

Next, assume for contradiction that the expected value of end-of-school skills across tracks $\mathbb{E}[\theta_2]$ is maximized under a track allocation mechanism that does not feature a skill threshold. Suppose that $P_V^* < P_A^*$ without loss of generality. By Lemma 1, these are the optimal instruction paces for the average skill level in track V and A, respectively. Therefore, $\mathbb{E}(\theta_1|S = V) < \mathbb{E}(\theta_1|S = A)$. Then, because there is no strict threshold, this means that for any initial skill level θ_1 there must be at least two children with initial skill levels smaller or equal than θ_1 that go to different tracks or at least two children with initial skill levels larger or equal than θ_1 that go to different tracks. This implies that there exists a child with $\theta_1' \leq \mathbb{E}(\theta_1|S=V)$ that goes to track $S=A$, and/or a child with $\theta_1' \geq \mathbb{E}(\theta_1|S=A)$ that goes to track $S=V$, and/or two children with skills $\theta_1' < \theta_1''$, with $\theta_1', \theta_1'' \in [\mathbb{E}(\theta_1|S=V), \mathbb{E}(\theta_1|S=A)]$, where the child with the smaller skill level goes to track A and the child with the larger skill level to track V.

However, given the condition in (27), this child with θ_1' would always benefit from being in the other track as the distance between her skill level and the average average skill level in that track is smaller than in the track she is in. Note, that moving just one child to another track does not change the average skills in both tracks. Thus, the policymaker can improve aggregate end-of-school skills by moving this child.

The same line of argument holds in the implied game that parents play when they endogenously sort their children into two tracks. If no skill threshold level exists, there is always a child that would unilaterally gain if sent to a different track.

Thus, we have established that the existence of a skill threshold is necessary for optimal end-of-school skills both if a policymaker makes the track allocation directly, and when parents play a sorting game. Next, we characterize the thresholds for both cases. Let $\tilde{\theta}_1$ be the skill threshold and let S again indicate to which track a child is allocated, now with $S=V$ for all $\theta_1 \leq \tilde{\theta}_1$ and $S=A$ for all $\theta_1 > \tilde{\theta}_1$.

A policymaker solves

$$\begin{aligned} & \max_{\tilde{\theta}_1} \quad \mathbb{E}(\theta_2) \\ \iff & \max_{\tilde{\theta}_1} \quad \mathbb{E}(\mathbb{E}(\theta_2|S)) \\ & \text{subject to} \\ & \qquad P_S \text{ chosen optimally given Lemma 1.} \end{aligned} \tag{28}$$

Using (26) and the law of iterated expectations, this maximization problem boils down to

$$\begin{aligned} & \max_{\tilde{\theta}_1} \quad \frac{\beta^2}{2\delta} + \frac{\gamma^2}{2\delta}\mathbb{E}\left(\mathbb{E}(\theta_1|S)^2\right) \\ \iff & \max_{\tilde{\theta}_1} \quad \frac{\beta^2}{2\delta} + \frac{\gamma^2}{2\delta}\left(F(\tilde{\theta}_1)\,\mathbb{E}(\theta_1|\theta_1 \leq \tilde{\theta}_1)^2 + (1-F(\tilde{\theta}_1))\,\mathbb{E}(\theta_1|\theta_1 > \tilde{\theta}_1)^2\right), \end{aligned} \tag{29}$$

where $F(.)$ denotes the cumulative distribution function of the normal distribution. Note that the right term is just the expected value (across tracks) of the conditional expected values of initial skills squared, conditional on the school track. This corresponds to the variance of the conditional expected values, which depend on the skill threshold $\tilde{\theta}_1$. Using the law of total variance, the maximization problem can thus be rewritten as (dropping the constant term)

$$\begin{aligned} & \max_{\tilde{\theta}_1} \quad \mathbb{E}(\theta_2) \\ \iff & \max_{\tilde{\theta}_1} \quad \frac{\gamma^2}{2\delta}\left(\sigma^2_{\theta_1} - \mathbb{E}(Var[\theta_1|S])\right). \end{aligned} \tag{30}$$

Thus, the policymaker chooses optimally a threshold such that the expected variance of skills in each track is minimized. The unique solution is then to set $\tilde{\theta}_1^* = \mathbb{E}\,\theta_1 = 0$, that is to split the distribution exactly in half. This makes the peer groups in each track as homogeneous as possible, which maximizes average and aggregate learning.

Next, we characterize the threshold that arises endogenously from the sorting game played by the parents. The equilibrium condition maintains that at this threshold, a parent is just indifferent between tracks as her child's skills would equivalently in both tracks. A parent of a child with skill $\hat{\theta}_1$ is indifferent between tracks V and A iff

$$\begin{aligned} & \left(\alpha + \hat{\theta}_1\frac{\gamma^2}{\delta}\right)\mathbb{E}(\theta_1|\theta_1 \leq \hat{\theta}_1) - \frac{\gamma^2}{2\delta}\,\mathbb{E}(\theta_1|\theta_1 \leq \hat{\theta}_1)^2 \\ = & \left(\alpha + \hat{\theta}_1\frac{\gamma^2}{\delta}\right)\mathbb{E}(\theta_1|\theta_1 > \hat{\theta}_1) - \frac{\gamma^2}{2\delta}\,\mathbb{E}(\theta_1|\theta_1 > \hat{\theta}_1)^2 \\ \iff & \left(-\alpha - \hat{\theta}_1\frac{\gamma^2}{\delta}\right)\sigma_{\theta_1}\frac{f(\hat{\theta}_1/\sigma)}{F(\hat{\theta}_1/\sigma)} - \frac{\gamma^2}{2\delta}\sigma^2_{\theta_1}\frac{f(\hat{\theta}_1/\sigma)^2}{F(\hat{\theta}_1/\sigma)^2} \\ = & \left(\alpha + \hat{\theta}_1\frac{\gamma^2}{\delta}\right)\sigma_{\theta_1}\frac{f(\hat{\theta}_1/\sigma)}{1-F(\hat{\theta}_1/\sigma)} - \frac{\gamma^2}{2\delta}\sigma^2_{\theta_1}\frac{f(\hat{\theta}_1/\sigma)^2}{(1-F(\hat{\theta}_1/\sigma))^2} \end{aligned} \tag{31}$$

where $F()$ denotes the CDF of a standard normally distributed random variable, and $f()$ is its density function. We solve for the root $\hat{\theta}_1$ that solves (31) numerically. In all cases with reasonable parameter values, (31) is a monotone function, such that the root is unique, if it exists. In the special case without direct peer externality, i.e. $\alpha = 0$, the solution is $\hat{\theta}_1 = 0$, as can be directly seen from (31). When $\alpha > 0$, the root is smaller than 0, i.e. $\hat{\theta}_1 < 0$.

Proposition 2

The proof of this Proposition follows directly from (A.1). In a comprehensive system the variance of initial skills across tracks is just equal to the overall variance since there is only one track. In a tracking system, the expected value of the conditional variances of skills across tracks is smaller than the overall variance, by the law of total variance and provided that the instruction paces are different across tracks. This holds for every skill threshold, not just for the optimal one. Thus average learning is higher.

To characterize the variance of θ_2, we start by collecting expressions for conditional and unconditional first and second moments.

The unconditional expected value of θ_2 in track V, if everyone went to V is

$$\begin{aligned}\mathbb{E}(\theta_{2,V}) &= \frac{\beta^2}{2\delta} + \alpha\bar{\theta}_{1,V} - \frac{\gamma^2}{2\delta}\bar{\theta}_{1,V}^2 \\ &= \frac{\beta^2}{2\delta} - \alpha\sigma_{\theta_1}\frac{f(\tilde{\theta}_1/\sigma_{\theta_1})}{F(\tilde{\theta}_1/\sigma_{\theta_1})} - \frac{\gamma^2}{2\delta}\sigma_{\theta_1}^2\frac{f(\tilde{\theta}_1/\sigma_{\theta_1})^2}{F(\tilde{\theta}_1/\sigma_{\theta_1})^2}.\end{aligned} \tag{32}$$

The unconditional expected value of θ_2 in track A, if everyone went to A is

$$\begin{aligned}\mathbb{E}(\theta_{2,A}) &= \frac{\beta^2}{2\delta} + \alpha\bar{\theta}_{1,A} - \frac{\gamma^2}{2\delta}\bar{\theta}_{1,A}^2 \\ &= \frac{\beta^2}{2\delta} + \alpha\sigma_{\theta_1}\frac{f(\tilde{\theta}_1/\sigma_{\theta_1})}{1-F(\tilde{\theta}_1/\sigma_{\theta_1})} - \frac{\gamma^2}{2\delta}\sigma_{\theta_1}^2\frac{f(\tilde{\theta}_1/\sigma_{\theta_1})^2}{(1-F(\tilde{\theta}_1/\sigma_{\theta_1}))^2}.\end{aligned} \tag{33}$$

The variance of θ_2 in a comprehensive system is

$$\begin{aligned}Var(\theta_{2,C}) &= \mathbb{E}(\big(\theta_2 - \mathbb{E}(\theta_2)\big)^2) \\ &= (1+\frac{\beta\gamma}{\delta})^2\sigma_{\theta_1}^2 + \sigma_{\eta_2}^2 \\ &\sigma_{\theta_2,C}^2 + \sigma_{\eta_2}^2,\end{aligned} \tag{34}$$

where we define $\sigma_{\theta_2,C}^2$ to be the variance of θ_2 net of the additive skill shock variance.

The unconditional variance of θ_2 in the V-track, that is the variance as if everyone went to V is

$$
\begin{aligned}
Var(\theta_{2,V}) &= \mathbb{E}\left(\theta_2 - \mathbb{E}(\theta_{2,V})\right)^2 \\
&= \mathbb{E}\left(\theta_1 + \alpha\bar{\theta}_{1,V} + \frac{\beta^2}{2\delta} + \frac{\beta\gamma}{\delta}\theta_1 + \frac{\gamma^2\theta_1\bar{\theta}_{1,V}}{\delta} - \frac{\gamma^2\bar{\theta}_{1,V}^2}{2\delta} + \eta_2 - (\alpha\bar{\theta}_{1,V} + \frac{\beta^2}{2\delta} - \frac{\gamma^2\bar{\theta}_{1,V}^2}{2\delta})\right)^2 \\
&= \sigma_{\theta_1}^2\left[(1+\frac{\beta\gamma}{\delta})^2 + 2(1+\frac{\beta\gamma}{\delta})\frac{\gamma^2}{\delta}\bar{\theta}_{1,V} + \frac{\gamma^4}{\delta^2}\bar{\theta}_{1,V}^2\right] + \sigma_{\eta_2}^2 \\
&= \sigma_{\theta_2,C}^2 + \sigma_{\theta_1}^2\left[2(1+\frac{\beta\gamma}{\delta})\frac{\gamma^2}{\delta}\bar{\theta}_{1,V} + \frac{\gamma^4}{\delta^2}\bar{\theta}_{1,V}^2\right] + \sigma_{\eta_2}^2 = \sigma_{\theta_{2,V}}^2 + \sigma_{\eta_2}^2
\end{aligned}
\tag{35}
$$

and similarly for track A

$$
Var(\theta_{2,A}) = \sigma_{\theta_2,C}^2 + \sigma_{\theta_1}^2\left[2(1+\frac{\beta\gamma}{\delta})\frac{\gamma^2}{\delta}\bar{\theta}_{1,A} + \frac{\gamma^4}{\delta^2}\bar{\theta}_{1,A}^2\right] + \sigma_{\eta_2}^2 = \sigma_{\theta_{2,A}}^2 + \sigma_{\eta_2}^2, \tag{36}
$$

where, again we define $\sigma_{\theta_{2,S}}^2$ to be the unconditional variance in track S net of the skill shock variance.

Given this, the conditional expected value of θ_2 in track S, that is the variance among those actually in that track is a truncation of the unconditional normally distributed $\theta_{2,S}$, where the truncation occurs at the skill level that a child with the initial cutoff skill $\tilde{\theta}_1$ obtained absent any shocks η_2, call it $\tilde{\theta}_2$. This skill level is just equal to the average unconditional skill $\theta_{2,S}$ as expressed above. Thus, we can find the conditional expected value of skills in track S using the formula for a truncated normal distribution and the computed unconditional standard deviation of θ_2 in each track,, $\sigma_{\theta_{2,S}}$, *as if* there were no skill shock realizations which yields

$$
\begin{aligned}
\mathbb{E}(\theta_{2,V}|S=V) &= \mathbb{E}(\theta_{2,V}) - \sigma_{2,V}\frac{f(\tilde{\theta}_1/\sigma_{\theta_2,V})}{F(\tilde{\theta}_1/\sigma_{\theta_2,V})} \\
&= \frac{\beta^2}{2\delta} + \alpha\bar{\theta}_{1,V} - \frac{\gamma^2}{2\delta}\bar{\theta}_{1,V}^2 - \sigma_{\theta_2,V}\frac{f(\tilde{\theta}_1/\sigma_{\theta_2,V})}{F(\tilde{\theta}_1/\sigma_{\theta_2,V})},
\end{aligned}
\tag{37}
$$

for the V track and

$$\begin{aligned}\mathbb{E}(\theta_{2,A}|S=A) &= \mathbb{E}(\theta_{2,A}) + \sigma_{2,A}\frac{f(\tilde{\theta}_1/\sigma_{2,A})}{1-F(\tilde{\theta}_1/\sigma_{\theta_2,A})} \\ &= \frac{\beta^2}{2\delta} + \alpha\bar{\theta}_{1,A} - \frac{\gamma^2}{2\delta}\bar{\theta}^2_{1,A} + \sigma_{\theta_2,A}\frac{f(\tilde{\theta}_1/\sigma_{\theta_2,A})}{1-F(\tilde{\theta}_1/\sigma_{\theta_2,A})},\end{aligned} \tag{38}$$

for the A track.

By the same logic but *adding* the skill shock variance $\sigma^2_{\eta 2}$, we can compute the conditional variances of θ_2 in each track as

$$Var(\theta_{2,V}|S=V) = \sigma^2_{\theta_2,V}\left(1 - \frac{\tilde{\theta}_1}{\sigma_{\theta_2,V}}\frac{f(\tilde{\theta}_1/\sigma_{\theta_2,V})}{F(\tilde{\theta}_1/\sigma_{\theta_2,V})} - \frac{f(\tilde{\theta}_1/\sigma_{\theta_2,V})^2}{F(\tilde{\theta}_1/\sigma_{\theta_2,V})^2}\right) + \sigma^2_{\eta 2},$$

$$Var(\theta_{2,A}|S=A) = \sigma^2_{\theta_2,A}\left(1 + \frac{\tilde{\theta}_1}{\sigma_{\theta_2,A}}\frac{f(\tilde{\theta}_1/\sigma_{\theta_2,A})}{1-F(\tilde{\theta}_1/\sigma_{\theta_2,A})} - \frac{f(\tilde{\theta}_1/\sigma_{\theta_2,A})^2}{(1-F(\tilde{\theta}_1/\sigma_{\theta_2,A}))^2}\right) + \sigma^2_{\eta 2}.$$

To obtain the variance of θ_2 in a tracking system, we make use of the law of total variance again

$$Var(\theta_{2,T}) = \mathbb{E}(Var(\theta_{2,S}|S)) + Var(\mathbb{E}(\theta_{2,S}|S)). \tag{39}$$

Under a general skill threshold level $\tilde{\theta}_1$, this yields a complicated expression for the unconditional variance. We therefore consider the special case of $\tilde{\theta}_1 = 0$, i.e. the case in which the policymaker makes the optimal track allocation choice. In that case, the expected value of the conditional variance of θ_2 across tracks can be simplified to

$$\mathbb{E}(Var(\theta_{2,S}|S)) = \sigma^2_{\eta 2} + (1 - \frac{f(0)^2}{0.25})\left(\sigma^2_{\theta_2,C} + \sigma^4_{\theta_1}\frac{\gamma^4}{\delta^2}\frac{f(0)^2}{0.25}\right). \tag{40}$$

Moreover, the variance of the conditional means of θ_2 across tracks is

$$\begin{aligned}Var(\mathbb{E}(\theta_{2,S}|S)) &= \left(\alpha\bar{\theta}_{1,V} - f(0)(\sigma_{2,v} + \sigma_{2,A})\right)^2 \\ &= \left(\alpha\bar{\theta}_{1,A} + f(0)(\sigma_{2,v} + \sigma_{2,A})\right)^2.\end{aligned} \tag{41}$$

Notice that the average squared distance to the overall mean of θ_2 is the same in both tracks. Collecting terms and simplifying gives the following expression for the unconditional variance of θ_2 in an optimal tracking system with $\tilde{\theta}_1 = 0$

$$
\begin{aligned}
Var(\theta_{2,T}) &= \sigma^2_{\eta_2} + (1-\chi)\sigma^2_{\theta_{2,C}} + \chi\frac{\gamma^4}{\delta^2}(2-4\chi)\sigma^4_{\theta_1} \\
&+ 2\chi\alpha\sigma_{\theta_1}(\alpha\sigma_{\theta_1} + \sigma_{\theta_{2,V}} + \sigma_{\theta_{2,A}}) + \chi\sigma_{\theta_{2,V}}\sigma_{\theta_{2,A}} \\
&= \sigma^2_{\eta_2} + \sigma^2_{\theta_{2,C}} + 2\chi\sigma^2_{\theta_1}\left(\alpha^2 + 2\alpha(1+\frac{\beta\gamma}{\delta}) + \frac{\gamma^4}{\delta^2}\sigma^2_{\theta_1}\right) \\
&- \left(4\chi\frac{\gamma^2}{\delta}\sigma^2_{\theta_1}\right)^2,
\end{aligned}
\tag{42}
$$

where $\chi \equiv 2f(0)^2 = \frac{1}{\pi}$. The condition that governs if the variance of end-of-school skills are larger than in a full tracking than in a full comprehensive system then reads

$$
\begin{aligned}
& Var(\theta_{2,T}) - Var(\theta_{2,C}) \geq 0 \\
\iff \quad & \alpha^2 + 2\alpha\left(1+\frac{\beta\gamma}{\delta}\right) - (8-\pi)\frac{\gamma^4}{\pi\delta^2}\sigma^2_{\theta_1} \geq 0.
\end{aligned}
\tag{43}
$$

This condition increasing monotonically in both α and β. Moreover it decreases in the variance of initial skills, $\sigma^2_{\theta_1}$.

Next, we show that a full tracking system leads to a "fatter" right tail of the end-of-school skill distribution compared to a comprehensive system. To see this, consider the child who, in expectation, has the highest end-of-school skill in a comprehensive system. Since θ_2 is monotonically increasing in θ_1 in a given track (see (26)), this is the child with the highest initial skill, say $\theta_{1,max}$. Moreover, from the properties of a truncated normal distribution, we know that, for any skill threshold $\tilde{\theta}_1$, average skills in the A track, $\bar{\theta}_{1,A}$ are larger than the unconditional average, $\bar{\theta}_{1,C} = 0$. Thus, the squared distance between $\theta_{1,max}$ and $\bar{\theta}_{1,A}$ in a tracking system is smaller. Taken together, (27) implies that the child with initial skill $\theta_{1,max}$ ends up with larger end-of-school skills compared to a comprehensive system, which skews the distribution positively.

Finally we derive the range of winners and loser from a tracking system relative to a comprehensive system. Given that θ_2 are monotonically increasing in θ_1 in every track, the range is characterized by the intersection of the linear function $\theta_{2,C}(\theta_1, \bar{\theta}_{1,C})$ with $\theta_{2,V}(\theta_1, \bar{\theta}_{1,V})$

and $\theta_{2,A}(\theta_1, \bar{\theta}_{1,A})$. For any skill threshold, the lower intersection $\theta_{1,L}$ hence solves

$$\begin{aligned} & \theta_{1,L} + \alpha\bar{\theta}_{1,C} + \frac{\beta^2}{2\delta} + \frac{\beta\gamma}{\delta}\theta_{1,L} + \frac{\gamma^2}{\delta}\bar{\theta}_{1,C}\theta_{1,L} - \frac{\gamma^2}{2\delta}\bar{\theta}^2_{1,C} + \eta_2 \\ = \; & \theta_{1,L} + \alpha\bar{\theta}_{1,V} + \frac{\beta^2}{2\delta} + \frac{\beta\gamma}{\delta}\theta_{1,V} + \frac{\gamma^2}{\delta}\bar{\theta}_{1,V}\theta_{1,L} - \frac{\gamma^2}{2\delta}\bar{\theta}^2_{1,V} + \eta_2 \\ \iff \quad \theta_{1,L} = \; & \frac{1}{2}\bar{\theta}_{1,V} - \frac{\alpha\delta}{\gamma^2}. \end{aligned} \tag{44}$$

Similarly, the upper intersection is given at

$$\theta_{1,U} = \frac{1}{2}\bar{\theta}_{1,A} - \frac{\alpha\delta}{\gamma^2}. \tag{45}$$

For any skill threshold $\tilde{\theta}_1$, the interval $[\theta_{1,L}, \bar{\theta}_{1,U}]$ is non-empty. Hence, there are always children with initial skill levels inside this interval who lose in terms of end-of-school skills in a full tracking system relative to a comprehensive system. Every child outside of this interval gains relative to the comprehensive system.

With $\alpha = 0$, the tracking skill threshold is at $\tilde{\theta}_1 = 0$ even if parents endogenously sort their children. Hence, children with initial skills inside a symmetric interval around 0, $[\frac{1}{2}\bar{\theta}_{1,V}, \frac{1}{2}\bar{\theta}_{1,A}]$, lose relative to a comprehensive track, since $\bar{\theta}_{1,V} = -\bar{\theta}_{1,A}$ if $\tilde{\theta}_1 = 0$. The average loss of a child in this interval is equal to $\frac{\gamma^2}{2\delta}\bar{\theta}^2_{1,V} = \frac{\gamma^2}{2\delta}\bar{\theta}^2_{1,A}$.

If $\alpha > 0$, and the policymaker enforces the tracking skill threshold $\tilde{\theta}_1 = 0$, the losses from tracking are concentrated among children in the V track. To see this, note that every child with initial skill in the interval $[\theta_{1,L}, 0]$ is allocated into the V track but loses relative to a comprehensive system. Similarly, every child with an initial skill inside $[0, \theta_{1,U}]$ is allocated to track A but loses relative to a comprehensive system. With $\alpha > 0$, $|\theta_{1,U}| < |\theta_{1,L}|$ and hence, the range of children in the A track that lose is smaller. The interval $[0, \theta_{1,U}]$ may even be empty in which case only children in the V track lose from tracking.

Proposition 3

First, we can derive the expected value of end-of-school skills in the 2-period model in a late tracking system as

$$
\begin{aligned}
\mathbb{E}(\theta_{3,LT}) &= \mathbb{E}(\mathbb{E}(\theta_{3,LT}|S^2_{LT})) \\
&= \mathbb{E}(\theta_{2,LT}) + \frac{\beta^2}{2\delta} + (\alpha + \frac{\beta\gamma}{\delta})\,\mathbb{E}(\mathbb{E}(\theta_{2,LT}|S_{LT})) + \frac{\gamma^2}{2\delta}\,\mathbb{E}(\mathbb{E}(\theta_{2,LT}|S_{LT})^2) \\
&= (2 + \alpha + \frac{\beta\gamma}{\delta})\frac{\beta^2}{2\delta} + \frac{\gamma^2}{2\delta}[\sigma^{\theta_2,LT} - \mathbb{E}(Var(\theta_{2,LT}|S_{LT}))],
\end{aligned} \tag{46}
$$

where $\mathbb{E}(\theta_{2,LT})$ and $\sigma^2_{\theta_2,LT}$ are just equal to the mean and variance of the comprehensive system in the one-period model (see equation (34)). The variable S_{LT} indicates the track selection in period 2, which follows the cut-off rule $S_{LT} = V$ if $\theta_{2,LT} \leq \tilde{\theta}_{2,LT}$ and $S_{LT} = A$ otherwise. The cut-off that maximizes (46) is $\tilde{\theta}^*_{2,LT} = \mathbb{E}(\theta_{2,LT}) = \frac{\beta^2}{2\delta}$. This follows as (46) mirrors that of average end-of-school skills in the full tracking system of the one-period model in that average and aggregate $\theta_{3,LT}$ decrease in the expected variance of skills in period 2 across tracks.

Similarly, we find the expected value of end-of-school skills in the 2-period model in an early tracking system as

$$
\begin{aligned}
\mathbb{E}(\theta_{3,ET}) &= \mathbb{E}(\mathbb{E}(\theta_{3,ET}|S^2_{ET})) \\
&= \frac{\beta^2}{2\delta} + (1 + \alpha + \frac{\beta\gamma}{\delta})\,\mathbb{E}(\mathbb{E}(\theta_{2,ET}|S_{ET})) + \frac{\gamma^2}{2\delta}\,\mathbb{E}(\mathbb{E}(\theta_{2,ET}|S_{ET})^2) \\
&= \frac{\beta^2}{2\delta} + (1 + \alpha + \frac{\beta\gamma}{\delta})\left(\frac{\beta^2}{2\delta} + \frac{\gamma^2}{2\delta}[\sigma^2_{\theta_1} - \mathbb{E}(Var(\theta_{1,ET}|S_{ET}))]\right) + \frac{\gamma^2}{2\delta}\,\mathbb{E}(\mathbb{E}(\theta_{2,ET}|S_{ET})^2 \\
&= \frac{\beta^2}{2\delta} + (1 + \alpha + \frac{\beta\gamma}{\delta})\left(\frac{\beta^2}{2\delta} + \frac{\gamma^2}{2\delta}[\sigma^2_{\theta_1} - \mathbb{E}(Var(\theta_{1,ET}|S_{ET}))]\right) \\
&+ \frac{\gamma^2}{2\delta}[\sigma^2_{\theta_2,ET} - \mathbb{E}(Var(\theta_{2,ET}|S_{ET}))].
\end{aligned} \tag{47}
$$

Comparing (46) and (47), the condition that governs if average end-of-school skills in a late tracking system are larger than in an early tracking system reads

$$
\begin{aligned}
&\mathbb{E}(\theta_{3,LT}) - \mathbb{E}(\theta_{3,ET}) \\
&= \frac{\gamma^2}{2\delta}\left(\mathbb{E}(\mathbb{E}(\theta_{2,LT}|S_{LT})^2) - \mathbb{E}(\mathbb{E}(\theta_{2,ET}|S_{ET})^2)\right) \\
&- (1 + \alpha + \frac{\beta\gamma}{\delta})\frac{\gamma^2}{\delta}\,\mathbb{E}(\mathbb{E}(\theta_1|S_{ET})^2) > 0.
\end{aligned} \tag{48}
$$

The last term of (48) represents the advantage of early tracking in the first stage of the schooling years. It stems from the smaller expected conditional variances of initial skills among children that are tracked relative to the overall variance. The conditional expected value of θ_2 in a late tracking system are given by

$$\mathbb{E}(\theta_{2,LT}|S_{LT}=V)=\frac{\beta^2}{2\delta}-\sigma_{\theta_2,LT}\frac{f(\tilde{\theta}_{2,LT}/\sigma_{\theta_2,LT})}{F(\tilde{\theta}_{2,LT}/\sigma_{\theta_2,LT})} \tag{49}$$

and

$$\mathbb{E}(\theta_{2,LT}|S_{LT}=A)=\frac{\beta^2}{2\delta}+\sigma_{\theta_2,LT}\frac{f(\tilde{\theta}_{2,LT}/\sigma_{\theta_2,LT})}{1-F(\tilde{\theta}_{2,LT}/\sigma_{\theta_2,LT})}, \tag{50}$$

where the unconditional variance of θ_2 in a late tracking system is given by $\sigma^2_{\theta_2,LT}=\sigma^2_{\theta_2,C}+\sigma^2_{\eta_2}$, i.e. by the one computed in equation (34). Since late tracking occurs *after* the realization of skill shocks in period 2, this variance additively *includes* the variance of these shocks. In contrast, the conditional expected values of θ_2 in an early tracking system, which are computed in equations (37) and (38) depend on $\sigma^2_{\theta_{2,S}}$, which as defined in (35) and (36) do not depend on the skill shock variance.

Condition (48) is generally ambiguous and hard to interpret for arbitrary skill thresholds. We focus again on the optimal tracking case, that is the case with skill threshold $\tilde{\theta}_1=\mathbb{E}(\theta_1)=0$ and $\tilde{\theta}_2=\mathbb{E}(\theta_{2,LT})=\frac{\beta^2}{2\delta}$. In that case, we can write the expressions for the various expected square conditional expected values as follows:

$$\begin{aligned}
\mathbb{E}(\mathbb{E}(\theta_1|S_{ET})^2)&=2\chi\sigma^2_{\theta_1}\\
\mathbb{E}(\mathbb{E}(\theta_{2,LT}|S_{LT})^2)&=\frac{\beta^4}{4\delta^2}+2\chi(\sigma^2_{\theta_{2,LT}}+\sigma^2_{\eta_2})\\
\mathbb{E}(\mathbb{E}(\theta_{2,ET}|S_{ET})^2)&=\frac{\beta^4}{4\delta^2}+2\chi\sigma^2_{\theta_1}\left(\alpha^2+\frac{\gamma^4}{\delta^2}f(0)^2\sigma^2_{\theta_1}-\frac{\beta^2\gamma^2}{2\delta^2}\right)\\
+2f(0)\sigma^2_{\theta_1}\left(\frac{\beta^2\gamma^2}{\delta^2}+2\alpha(1+\frac{\beta\gamma}{\delta})-(2\frac{\gamma^2}{\delta}f(0)\sigma_{\theta_1})^2\right)&+2\chi(\sigma^2_{\theta_{2,LT}}+2\chi\frac{\gamma^4}{\delta^2}\sigma^2_{\theta_1}).
\end{aligned}$$

Condition (48) then becomes

$$
\begin{aligned}
&\mathbb{E}(\theta_{3,LT}) - \mathbb{E}(\theta_{3,ET}) \\
&= \frac{\gamma^2}{2\delta}\Bigg(2\chi\sigma^2_{\eta_2} - 2\chi\sigma^2_{\theta_1}\Bigg(\alpha^2 + \frac{\gamma^4}{\delta^2}f(0)^2\sigma^2_{\theta_1} - \frac{\beta^2\gamma^2}{2\delta^2} \\
&+ \frac{\beta^2\gamma^2}{\delta^2} + 2\alpha(1+\frac{\beta\gamma}{\delta}) - 4\frac{\gamma^4}{\delta^2}f(0)^2\sigma^2_{\theta_1} + 2\chi\frac{\gamma^4}{\delta^2}\sigma^2_{\theta_1} + 1 + \alpha + \frac{\beta\gamma}{\delta}\Bigg)\Bigg) \\
&= \frac{\gamma^2}{\pi\delta}\Bigg(\sigma^2_{\eta_2} - \sigma^2_{\theta_1}\Bigg(1 + \alpha + \alpha^2 + \frac{\beta\gamma}{\delta} + \frac{\beta^2\gamma^2}{2\delta^2} + 2\alpha(1+\frac{\beta\gamma}{\delta}) + \frac{\gamma^4}{2\delta^2\pi}\sigma^2_{\theta_1}\Bigg)\Bigg) > 0.
\end{aligned}
\tag{51}
$$

From this, Proposition 3 follows.

A.2 Equilibrium Definition

We introduce some notation to define the equilibrium more easily. Let $x_j \in\in X_j$ be the age-specific state vector of an individual of age j, as defined by the recursive representation of the individual's problems in Section 3. Let its stationary distribution be $\Theta(X)$. Then, a stationary recursive competitive equilibrium for this economy is a collection of: (i) decision rules for college graduation $\{d^E(x_5)\}$, for school track $\{d^{S'}(x_{11})\}$, consumption, labor supply, and assets holdings $\{c_j(x_j), n_j(x_j), a_j(x_j)\}$, and parental transfers $\{a'_5\,(x_j)\}$; value functions $\{V_j\,(x_j)\}$; (iii) aggregate capital and labor inputs $\{K, H_0, H_1\}$; (iv) prices $\{r, w^0, w^1\}$; and (v) average skill levels among children in school track S' $\{\bar{\theta}'_{j',S'}\}$ such that:

1. Given prices and average skill levels among children in each school track, decision rules solve the respective household problems and $\{V_j\,(x_j)\}$ are the associated value functions.

2. Given prices, aggregate capital and labor inputs solve the representative firm's problem, i.e. it equates marginal products to prices.

3. Given average skill levels among children in each school track, allocation of children in school track solves the parent's problem, i.e. actual average skill levels are consistent with parents' prior.

4. Labor market for each education level clears.
   For high-school level:

$$H_0 = \sum_{j=5}^{16} \int_{X_j} n_j(x_j)\; h_j\,(x_j)\, d\Theta(X \mid E=0) + \sum_{j=5}^{5} \int_{X_j} n_j(x_j)\; h_j\,(x_j)\, d\Theta(X \mid E=1)$$

where the first summation is the supply of high-school graduates while the second is the labor supply of college students.

For college level:

$$H_1 = \sum_{j=6}^{16} \int_{X_j} n_j(x_j)\; h_j\,(x_j)\, d\Theta(X \mid E=1).$$

5. Asset market clears

$$K = \sum_{j=5}^{20} \int_{X_j} a_j(x_j) d\Theta(X),$$

which implies that the goods market clears;

6. The distribution of X is stationary: $\Theta(X) = \int \Gamma(X) d\Theta(X)$.

A.3 German Education System

In this section, we provide an overview about the most important features of the German Education and School System. A more extensive description can be found, for example, in Henninges et al. (2019). Figure S1 illustrates a simplified structure of the system, starting in Grade 4 and ending with tertiary education.

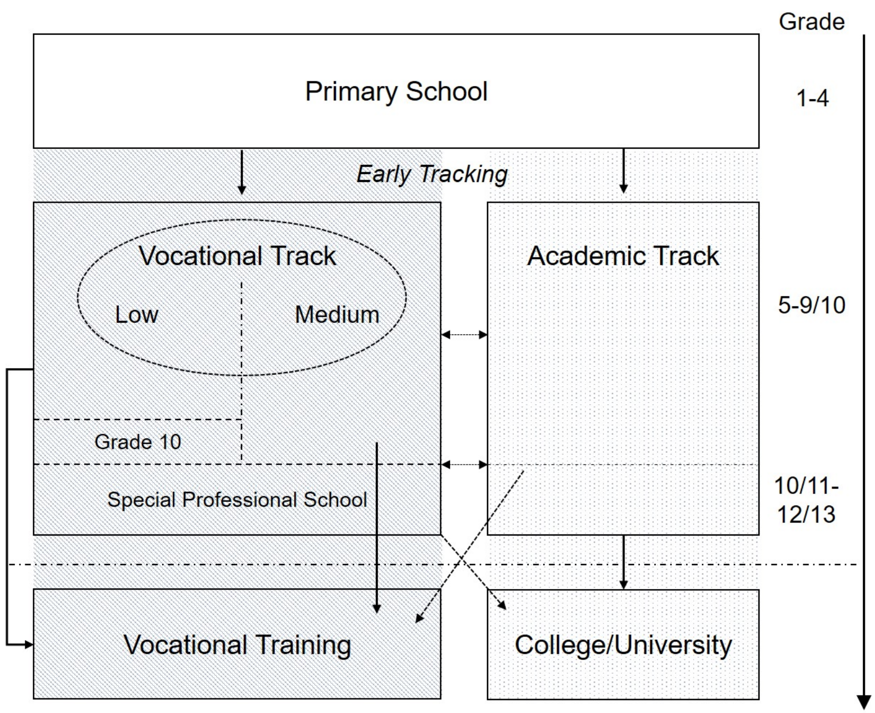
Generally, schooling is mandatory in Germany for every child starting at age six and lasting for nine or ten years. At age six, all children visit a comprehensive primary school that last the first four grades.[63] After that, children are allocated into traditionally three different secondary school tracks: A lower vocational track, a medium vocational track and an academic track. However, triggered by the so-called PISA shock in the early 2000s, federal states in Germany have started reforming their secondary school system. In particular, the two vocational tracks have often been combined into one, resulting in a two-track system in the majority of federal states (Bellenberg and Forell, 2012). For that reason, and because even if still two vocational tracks exist, they are much more similar in comparison to the academic track schools, we opt to restrict our analysis in this paper to two school tracks.

Generally, the school tracks differ in the curricula taught, the length of study, and the end-of-school qualifications that come with graduation. In particular, only the academic track schools end with a university entrance qualification that directly allows children to go

[63] In two federal states, Berlin and Brandenburg, comprehensive primary school lasts the first 6 grades.

to college. This requires the completion of the second stage of secondary school, typically grades 10/11 to 12/13. Graduating from a vocational track occurs after grade 9 and 10 and allows children to take up vocational training in blue-collar jobs or proceed to a professional school that prepares for entry into white-collar, business or skilled trade occupations. At this stage, there is considerable scope for mobility between tracks. Firstly, professional degrees often allow access to university studies in selected fields. Secondly, children can directly switch to an academic track school if their school marks and achievements admit that. Finally, after having worked for a number of years in a vocational jobs, access to some college degrees can be possible. At the same time, it is of course possible to switch from an academic track school into a vocational training or job after the mandatory education has been completed.

Figure S1: Simplified Structure of the German Education System



The public expenditure per student does not differ significantly across school tracks. Table S1 lists average per-student expenditures across the various school types in the years 2010 to 2020. Across these years, public expenditures by student were highest in pure lower vocational track schools. Expenditures in academic track schools were roughly equal compared to expenditures in joint vocational track schools. The bulk of these expenditures is attributable to teacher pay (around 80%) and the rest for investments into buildings, equipment etc. This suggests that resource differences across school tracks should not be a main driver behind achievement differences, on average.

Table S1: Per-Student Public Expenditures across School Types and Years

Year	Primary	Lower Voc.	Upper Voc.	Joint Voc.	Acad.	Compr.
2010	5,200 €	7,100 €	5,300 €	8,000 €	6,600 €	6,600 €
2011	5,500 €	7,300 €	5,600 €	8,000 €	7,100 €	7,100 €
2012	5,400 €	7,900 €	5,700 €	7,700 €	7,200 €	7,200 €
2013	5,600 €	8,200 €	5,900 €	7,700 €	7,500 €	7,500 €
2014	5,900 €	8,700 €	6,200 €	8,000 €	7,800 €	7,800 €
2015	6,000 €	8,900 €	6,400 €	8,000 €	7,900 €	8,000 €
2016	6,200 €	9,300 €	6,700 €	8,100 €	8,100 €	8,200 €
2017	6,400 €	9,800 €	7,000 €	8,300 €	8,500 €	8,600 €
2018	6,700 €	10,400 €	7,400 €	8,700 €	8,800 €	9,100 €
2019	7,100 €	11,200 €	7,900 €	9,200 €	9,300 €	9,500 €
2020	7,400 €	12,200 €	8,200 €	9,500 €	9,600 €	10,000 €

Source: Statistisches Bundesamt (Bildungsfinanzbericht, Bildungsausgaben - Ausgaben je Schüler, Sonderauswertung)

A remaining driver behind achievement differences across school tracks could be the teaching quality. In particular, higher-quality teachers could select for academic track schools. However, regardless of the secondary school track, becoming a teacher requires university studies in the range of 7 to 10 semesters and a similar university degree. On top of that, the differences in wages across school tracks are no longer significant in many federal states. For example, both tenured teachers at vocational track schools and teachers at academic track schools are eligible for the same public pay grade in most northern and eastern federal states already.

A.4 Measuring Child Skills in the NEPS

In this section, we provide an overview of our measures of child skills. One of the main goals of the NEPS project is to document the development of competencies of individuals over their lifespan (Neumann et al., 2013). To that end, the NEPS carefully designs and implements regular tests of the respondents' competencies along several domains. Given its central role not only in educational contexts but also as a predictor for later labor market success, we focus on mathematical competencies. Following the guidelines set by the Program for International Student Assessment (PISA), the mathematical competence domain is not just designed to assess the extent to which children have learned the content of school curricula but also to judge a child's ability to use mathematics to constructively engage with real-life problems (Neumann et al., 2013). The test, therefore, includes items related to "overarching" mathematical content areas that are consistent across all ages, such as quantity, change & relationships, space & shape, as well as several cognitive components, such as mathematical communication, argumentation, or modeling. The age-specific test items include primarily simple and complex multiple-choice questions, as well as short-constructed responses.[64]

In order to use these questions for the analysis of latent competencies, they need to be scaled. The NEPS (similar to the PISA) uses a scaling procedure that follows item response theory (IRT). IRT is a popular instrument in psychometrics to extract latent ability or other factors from test data. To quote the NEPS: "IRT was chosen as scaling framework for the newly developed tests because it allows for an estimation of item parameters independent of the sample of persons and for an estimation of ability independent of the sample of items. With IRT it is possible to scale the ability of persons in different waves on the same scale, even when different tests were used at each measurement occasion" (Pohl and Carstensen, 2013).

The most important scaling model used by the NEPS is the Rasch model. This model assumes that the right answers given to a set of questions by a number of respondents contain all information needed to measure a person's latent ability as well as the question's difficulty. It does so by positing that the probabillity that person v gives the right answer to question i is given by:

$$p(X_{vi} = 1) = 1 - p(X_{vi} = 0) = \frac{\exp(\theta_v - \sigma_i)}{1 + \exp(\theta_v - \sigma_i)}, \tag{52}$$

where θ_v denotes the latent ability of person v and σ_i is a measure of the question's difficulty. Thus, this model maps the total sum score of an individual into an ability parameter estimate. The scale is arbitrary. However, the ability estimate is cardinal.[65] This model is estimated via (weighted) conditional maximum likelihood under a normality assumption on latent ability.

Table S2 describes NEPS samples of mathematics assessments by cohort and Grade level.

A.5 Details on Child Skill Technology Estimation

A.5.1 Skills Measurement

We employ a linear measurement system for the logarithm of latent skills in each period that is given by

$$M_{i,k,j} = \mu_{k,j} + \lambda_{k,j}\theta_{i,j} + \epsilon_{i,k,j}, \tag{53}$$

[64]A simple multiple choice question consists of one correct out of four answer categories, and complex multiple choice questions consist of a number of subtasks with one correct answer out of two options. Short-constructed responses typically ask for a number (Pohl and Carstensen, 2012). The mathematical competence test primarily consists of simple multiple-choice questions.

[65]It is interval-scaled asBallou (2009) puts it. That means an increase of 5 points from 15 to 20 represents the same gain in achievement as from 25 to 30.

Table S2: NEPS Mathematic Assessment Samples

			Information on Parents' Education		Information on School Track	
		Obs.	Obs.	% College Parents	Obs.	% Ac. Track
Cohort 1	K1	2,014	1,709	51%		
Cohort 2	G1	6,352	5,784	46%	2,731	63%
	G2	5,888	5,425	47%	2,651	62%
	G4	6,610	6,068	46%	3,229	63%
	G7	2,479	2,410	51%	2,208	58%
Cohort 3	G5	5,193	3,856	38%	4,369	52%
	G7	6,191	4,214	38%	5,525	49%
	G9	4,888	3,387	38%	4,356	47%
	G12*	3,785	2,830	41%	3,331	58%
Cohort 4	G9	14,523	8,474	35%	14,215	40%
	G12*	5,733	3,767	24%	5,530	23%

Notes: This table describes NEPS mathematics assessments by cohort. Note that in Grade 12 the assessments are different by school track which makes the comparison of test scores by parental education or school track impossible. Source: NEPS

where $M_{i,k,j}$ denotes the kth measure for latent log skills of child i in period j. In each period, we have at least 3 different measures in our data, which typically constitute the achievement (item response theory) test scores of each child and are discussed in detail below. The parameters $\mu_{k,j}$, and $\lambda_{k,j}$ denote the location and factor loading of latent log skills, respectively. By $\epsilon_{i,k,j}$, we denote the measurement error. The parameters and measures are defined conditional on child's age and gender, which we keep implicit.

Following Cunha et al. (2010), we normalize $\mathbb{E}(\theta_j) = 0$ and $\lambda_{1,j} = 1$ for all j. That is, the first-factor loading is normalized to 1 in all periods.[66] We further normalize the measurement errors, such that $E(\epsilon_{k,j}) = 0$ for all j. Given that, the location parameters $\mu_{k,j}$ are identified from the means of the measures. In order to identify the factor loadings, we further assume that the measurement errors are independent of each other across measures and independent from latent skills. Under these assumptions and given that we have at least three measures of latent skills available in each period, we can identify the loadings on each measure in each

[66] We are aware of the potential bias that can arise from this assumption (see Agostinelli and Wiswall (2016)). However, contrary to their case, we measure three different stages of child development, where each stage comes with a new cohort of children (see below). Thus we cannot follow children over multiple periods. Moreover, even if we could, the data we use does not contain age-invariant measures according to their definition.

period by ratios of covariances of the measures (as in Agostinelli et al. (2019)):

$$\lambda_{k,j} = \frac{Cov(M_{k,j}, M_{k',j})}{Cov(M_{1,j}, M_{k',j})} \tag{54}$$

for all $k, k' > 1$ and $k \neq k'$. Rescaling the measures by their identified location and scale parameters then gives us error-contaminated measures of latent skills for each period as

$$\theta_{i,j} = \frac{M_{i,k,j} - \mu_{k,j}}{\lambda_{k,j}} - \frac{\epsilon_{i,k,j}}{\lambda_{k,j}} = \widetilde{M}_{i,k,j} - \frac{\epsilon_{i,k,j}}{\lambda_{k,j}}. \tag{55}$$

Equipped with identified latent variables up to measurement error for all periods, we can plug these into the child skill technology (25), which yields

$$\begin{aligned} \widetilde{M}_{i,k,j+1} = \kappa_{0,j} + \kappa_{1,j}\widetilde{M}_{i,k,j} + \kappa_{2,j}\widetilde{M}^2_{i,k,j} + \kappa_{3,j}\overline{\widetilde{M}}_{-i,j,S} \\ + \kappa_{4,j}(\widetilde{M}_{i,k,j} - \overline{\widetilde{M}}_{j,S})^2 + \kappa_{5,j}E_i + \zeta_{i,k,j+1}, \end{aligned} \tag{56}$$

where $\overline{\widetilde{M}}_{-i,j,S}$ refers to the expected value of the kth transformed measure across all children other than i in a classroom in track S and $\overline{\widetilde{M}}_{j,S}$ to that of the expected value of the measures across all children in a school that belongs to track S.

Importantly, the residual $\zeta_{i,k,j+1}$ now contains not only structural skill shocks, $\eta_{i,j+1}$, but also the measurement errors, $\epsilon_{i,k,j}$ as well as interactions of the measurement error with the rescaled measures and even the variance of the measurement errors. For that reason, even if a standard assumption of mean independence of the structural shocks η conditional on all independent variables holds, an OLS estimator of (56) will be biased. To account for that, we follow the literature and use excluded measures as instrumental variables, which we describe in Appendix A.5.[67]

Tables S3 and S4 describe the evolution of child skills over time using the identified latent variables.

[67] Under the assumption that measurement error is uncorrelated across measures, this strategy will take of measurement error and the interaction terms included in (56) but not of the variances of the measurement error. These will show up in the estimated intercept, thus biasing the constant. Since this constant does not have an economic meaning in our model, we disregard this bias for now. In the future, we can recover the variance of the measurement errors using ratios of covariances of the measures again, as in Cunha et al. (2010).

Table S3: Differences in Average Skills in Standard Deviation

	Grade	Difference by Parent's Education	School Track
Cohort 2	G1	0.46	0.77
	G4	0.55	0.83
	G7	0.45	0.93
Cohort 3	G5	0.48	0.87
	G7	0.55	0.92
	G9	0.57	1.01
Cohort 4	G9	0.54	0.97

Notes: This table provides information on average differences identified latent math grades up to measurement error in one standard deviation unit by parental background and school track over time. All observations are weighted. Source: NEPS

Table S4: Rank-Rank Correlations

		Rank-Rank Correlation	Obs.
Cohort 2	G1-G4	0.57	3,116
	G1-G7	0.61	2,267
Cohort 3	G5-G9	0.67	4,927

Notes: This table provides the rank-rank correlation in identified latent math grades up to measurement error. Source: NEPS

A.5.2 Instrumental Variables and Data

We estimate (56) using the (IRT) measure of mathematics tests. This is because we have this measure available at every stage j. We consider three stages of the schooling career, corresponding to the timing of our model. The first stage is the second period in a child's life and therefore indexed by $j = 2$ and corresponds to 4 years of primary school in real life, where children are typically aged 6 to 10. To estimate the parameters of this stage, we use the NEPS Starting Cohort 2. To account for measurement error, we instrument the math test scores in the first grade of primary school using test scores on science and vocabulary in the first grade of primary school as well as a math test in the second grade of primary school.

The second stage corresponds to the third period in a child's life, $j = 3$, when they are between 10 and 14 years old and typically go to secondary school. To estimate the parameters, we use data from the NEPS Starting Cohort 3. This data set, unfortunately, contains a relatively low number of observations for the first two years of secondary school. For that reason, we estimate the child technology parameters in that stage once on math test scores between grades 5 and 9 and once on math test scores between grades 7 and 9, after an increase in the sample size. The instruments are reading test scores in grade 5 and science test scores in grade 6 in the former case, and reading and orthography test scores (both in grade 7) in the latter case.

The third stage corresponds to the fourth period in a child's life, $j = 4$ when they are between 14/15 and 18 years old and typically finish secondary school. We use the NEPS Starting Cohort 4 to estimate the technology parameters in this case, again relying on transformed math test scores in grade 9 and grade 12.[68] The instruments we employ are vocabulary, science, and reading test scores in grade 9.

For the primary school stage, we restrict the sample to children who are in classes with a size of at least 5 children such that we can compute a meaningful class average. In both secondary school stages, we restrict the class sizes to be at a minimum of 8 children. This is because it is not uncommon that some primary schools, especially in rural areas, feature

[68] In Germany, the vocational track schools typically end after grade 9 or grade 10 and so-called upper secondary schooling only happens in academic track schools. However, the NEPS data keeps track of the students even if they are no longer enrolled in a school and tests them at the same age. A remaining issue is, of course, that even though we know the classroom compositions in grade 9, we do not know how long learning in that classroom continues in a vocational track school. For that reason, we make the assumption that children who went to a vocational track school that finished before they are 18 years old, continue to learn in an environment that is the same as if the vocational school had continued. In reality, students who graduate from vocational schools often continue with an apprenticeship, where we think it reasonable to assume that the peer composition is similar to the one in school.

quite small class sizes in Germany. In contrast, class sizes are typically in the range of 20-30 in secondary school.

A.6 Supplementary Tables

Table S5: OLS Estimates using Class-specific direct Peer Effects

		Dependent Variable: $\theta_{i,j+1}$ in model period			
		$j=2$	$j=3$		$j=4$
		Age Sample			
Coefficient	Variable	6-10	10-14/15	12-14/15	14/15-18
$\hat{\kappa}_{1,j}$	$\theta_{i,j}$	0.577	0.564	0.587	0.522
		(0.016)	(0.025)	(0.024)	(0.021)
$\hat{\kappa}_{2,j}$	${\theta_{i,j}}^2$	-	0.162	0.189	0.137
		-	(0.115)	(0.110)	(0.070)
$\hat{\kappa}_{3,j}$	$\bar{\theta}_{-i,S,j}$	-0.020	0.145	0.196	0.054
		(0.047)	(0.099)	(0.081)	(0.068)
$\hat{\kappa}_{4,j}$	$(\theta_{i,j} - \bar{\theta}_{S,j})^2$	-0.088	-0.229	-0.304	-0.193
		(0.047)	(0.126)	(0.124)	(0.081)
$\hat{\kappa}_{5,j}$	$E=1$	0.046	0.019	0.023	0.017
		(0.007)	(0.009)	(0.008)	(0.006)
$\hat{\kappa}_0$	Constant	-0.020	-0.003	-0.007	-0.097
N Children		3,925	1,985	2,582	3,072
N Schools		339	137	188	241
R^2		0.455	0.5211	0.5815	0.5134

Models control for age, gender and school fixed effects.
Standard errors are clustered at the school level

Table S6: OLS Estimates using Track-specific direct Peer Effects

		Dependent Variable: $\theta_{i,j+1}$ in model period			
		$j=2$	$j=3$		$j=4$
		Age Sample			
Coefficient	Variable	6-10	10-14/15	12-14/15	14/15-18
$\hat{\kappa}_{1,j}$	$\theta_{i,j}$	0.577	0.558	0.598	0.521
		(0.016)	(0.025)	(0.024)	(0.021)
$\hat{\kappa}_{2,j}$	${\theta_{i,j}}^2$	-	0.174	0.197	0.142
		-	(0.115)	(0.110)	(0.070)
$\hat{\kappa}_{3,j}$	$\bar{\theta}_{-i,S,j}$	-	0.569	0.419	0.166
		-	(0.194)	(0.095)	(0.136)
$\hat{\kappa}_{4,j}$	$(\theta_{i,j} - \bar{\theta}_{S,j})^2$	-0.088	-0.240	-0.316	-0.199
		(0.047)	(0.127)	(0.122)	(0.081)
$\hat{\kappa}_{5,j}$	$E=1$	0.046	0.019	0.023	0.017
		(0.007)	(0.009)	(0.008)	(0.006)
$\hat{\kappa}_0$	Constant	-0.020	-0.013	-0.013	-0.107
N Children		3,926	1,985	2,582	3,072
N Schools		340	137	188	241
R^2		0.455	0.5211	0.5815	0.5134

Models control for age, gender and school fixed effects.

Standard errors are clustered at the school level

Table S7: IV Estimates using Track-specific direct Peer Effects

		Dependent Variable: $\theta_{i,j+1}$ in model period			
		$j=2$	$j=3$		$j=4$
		Age Sample			
Coefficient	Variable	6-10	10-14/15	12-14/15	14/15-18
$\hat{\kappa}_{1,j}$	$\theta_{i,j}$	0.956	1.132	0.899	0.778
		(0.027)	(0.136)	(0.062)	(0.055)
$\hat{\kappa}_{2,j}$	$\theta_{i,j}^2$	-	0.333	0.591	0.308
		-	(0.317)	(0.277)	(0.157)
$\hat{\kappa}_{3,j}$	$\bar{\theta}_{-i,S,j}$	-	-26.049	5.696	-0.422
			(27.718)	(6.825)	(4.080)
$\hat{\kappa}_{4,j}$	$(\theta_{i,j}-\bar{\theta}_{S,j})^2$	-0.211	-0.621	-0.921	-0.591
		(0.136)	(0.549)	(0.411)	(0.211)
$\hat{\kappa}_{5,j}$	$E=1$	0.022	0.004	0.010	0.008
		(0.007)	(0.016)	(0.008)	(0.007)
$\hat{\kappa}_0$	Constant	-0.046	2.446	0.409	-0.029
N Children		3,530	1,934	2,580	2,934
N Schools		327	137	188	240
R^2		0.364		0.440	0.448

Models control for age, gender and school fixed effects.

Standard errors are clustered at the school level