Efficiency and Equity of Education Tracking
A Quantitative Analysis

Suzanne Bellue Lukas Mahler*

February 2023

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 a 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 depend on the design of the instruction levels in each school track.

Keywords: Intergenerational Mobility, Education Tracking, Equality of Opportunity

JEL codes: E24, I24, J24

*Bellue: University of Mannheim (suzanne.bellue@uni-mannheim.de); Mahler: University of Mannheim (lukas.mahler@gess.uni-mannheim.de); We thank Antonio Ciccone, Minchul Yum, Michele Tertilt, Andreas Gulyas and participants at various conferences and seminars for useful comments and discussions. Financial support from the German Science Foundation through CRC TR 224 (Project A04) and the SFB 884 Political Economy of Reforms is gratefully acknowledged.

1. Introduction

School tracking – the allocation of school children into different types of schools at some point during their school career – is a common feature of education policy across OECD countries.[1] While school tracking is designed to improve teaching efficiency, it may hamper equality of opportunity in access to education. The argument behind tracking efficiency is that grouping children according to their ability and aspirations creates more homogeneous peer groups and allows for tailored instruction levels and curricula, which improve the educational outcomes of children (Bonesrønning et al., 2022; Duflo, Dupas, and Kremer, 2011). However, because the track decision is in reality often influenced by family background, tracking may impair mobility in educational and labor market outcomes across generations (Dustmann, 2004; Falk, Kosse, and Pinger, 2021; Meghir and Palme, 2005; Pekkala Kerr, Pekkarinen, and Uusitalo, 2013). The parental influence on the track decision seems particularly strong if tracking occurs at a relatively young age, when measures of child ability are especially noisy (Hanushek and Wössmann, 2006). For that reason, it is not surprising that school tracking, and in particular its timing, is a frequently recurrent issue in the public and academic debate about education reforms in countries with a strict and early tracking regime, such as Germany.[2]

In this paper, we contribute to that debate by providing a quantitative assessment of the long-run aggregate, distributional and inter-generational effects of school tracking policies. In light of the arguments above, any such assessment needs to take into account the effects of tracking on the educational outcomes of children, as well as how these outcomes translate into labor market outcomes and outcomes across generations. 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.[3] Macroeconomic models of mobility provide a useful environment to consider such effects, but have so far largely ignored how the development of child skills during school is affected through

[1] An overview about school tracking policies in OECD countries is given in Chapter 2 in OECD (2013b). We differentiate school tracking, which refers to the allocation of students into physically distinct types of schools that differ in the curriculum taught, intensity and length, from ability grouping within a school, where the curriculum and educational goals remain the same.

[2] There is substantial variation in the timing of 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 that US and UK do not track at all during secondary school.

[3] For example, if the share of children who are allocated to an academic track school increases substantially, in the long run, 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.

peers and teaching levels. We fill this gap by building a macroeconomic model of overlapping generations that explicitly zooms in on the schooling years of children.

The model is built 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 child skills directly through interactions with peers at her 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 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, it can be optimal to postpone tracking, even from an efficiency point of view. Moreover, we show that tracking can increase overall inequality in educational outcomes relative to a comprehensive school system. 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 and embeds the main arguments about school tracking that are frequently made in the public discourse.[4]

We then embed this child skill development theory into a full general equilibrium life-cycle Aiyagari framework of overlapping generations, in which parents care about their offspring in the tradition of Becker and Tomes (1986) and child skills during the school career evolve according to our technology 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

[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.

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 is solved numerically, and the parameters are calibrated in two steps. First, we estimate the parameters of the child skill formation technology directly from German data on school children using a latent variable framework as in Cunha, Heckman, and Schennach (2010). In particular, we use information on achievement test scores as measures for child skills at different stages of their school careers. To account for measurement error, we use an instrumental variable strategy similar to Agostinelli, Saharkhiz, and Wiswall (2019). 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 in 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 validity of the model 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, Puhani, and Schönberg (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 for lifetime inequality across the population, the school tracking policy plays an important role. In particular, variation in the initial school track alone can account for 15% of the variation in lifetime earnings and 19% of variation in lifetime wealth. As in the data, parental education is, after child skills, the second most important determinant of initial school track choice. We use our model to show that the majority of this effect comes from direct parental preferences for children to follow in their own 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 would optimally suggest a vocational track school. This harms not only her own child's learning outcomes but also affects average learning in that track as the instruction pace endogenously adjust 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 small (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, as a means to improve equality of opportunity in access to academic education (Woessmann, 2013). We show that postponing the tracking age indeed succeeds in improving social mobility as it leads to a 8-10% decrease in the intergenerational elasticity of income. However, this comes at a cost of around 3% of GDP. The reason for this significant drop is that postponing tracking incentivizes the majority of parents to send their children to academic track schools, regardless of their own education. This results in significant 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; Fuchs-Schündeln, Krueger, Ludwig, et al., 2022; Fujimoto, Lagakos, and Vanvuren, 2022; Jang and Yum, 2022;

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), and Fuchs-Schündeln, Krueger, Ludwig, et al. (2022) who model high-school graduation choice. However, all of these studies except Fujimoto, Lagakos, and Vanvuren (2022) 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. In that sense, our paper is perhaps most closely related to to Fujimoto, Lagakos, and Vanvuren (2022) who study the importance of free secondary schooling for misallocation driven by borrowing constraints in Ghana. Our contribution is to analyze the 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, Agostinelli, Doepke, et al. (2023), Agostinelli, Saharkhiz, and Wiswall (2019), Cunha and Heckman (2007), and Cunha, Heckman, and Schennach (2010)). Our main innovation relative to this literature is to consider 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, Dupas, and Kremer (2011)'s evidence in Kenyan primary schools, we consider how the instruction levels adjust endogenously to the peer composition in schools of a certain 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, Salvanes, and Vaage (2010), for Norway; Malamud and Pop-Eleches (2011), for Romania; Pekkala Kerr, Pekkarinen, and Uusitalo (2013), for Finland; and Matthewes (2021) and 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, Maurin, and McNally (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 who lost their most able students. A notable exception is Dustmann, Puhani, and Schönberg (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 mis-allocation 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 develops a simple theory for the child skills formation during her school years. Section 3 presents the full life-cycle Aiyagari GE framework of overlapping generations that incorporates this theory. Section 4 explains how we estimate and parameterize the model. It also offers some validation exercises. In Section 5, we use the calibrated model to perform a series of counterfactual experiments to quantify the effects of different school tracking policy regimes. Finally, Section 6 concludes.

2. A Model of Child Skill Formation during School Years

In this section, 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. We use the model to derive a series of theoretical implications about the effects of school tracking policies on the distribution of end-of-school skills that are in line with the most robust empirical findings concerning these effects. As the end-of-school skills form the basis of adult human capital and, thus, lifetime economic outcomes, the child skill formation technology constitutes the main cornerstone of our full quantitative model.

2.1. Child Skill Formation

We assume that each child i arrives just before entering school with a set of skills (or abilities) that can be summarized in a univariate level, $\theta_{i,j^c=1}$.[5] The index j^c denotes the stage of the schooling career, where each stage corresponds to a period (age group) during the school life of a child. 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.

The initial skill level $\theta_{i,j^c=1}$ is likely a function of genetic components and parental skills directly, but also 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). Given our focus on school tracking policies, we take the initial distribution of skills across children from different parental backgrounds upon (primary) school enrollment as given.

The schooling system in a given stage is characterized by the number of distinct school tracks. If there is only one track to which all schools belong, we speak of a *comprehensive system*. If there are multiple distinct school tracks, we speak of a *tracking system*. Every school, and hence every child must belong to one track, which we denote by S^c. The sorting mechanism of children across tracks can vary and is discussed in detail below. School tracks differ ex-ante only in their *pace of instruction*, denoted by P_{S^c}. We think of the pace of instruction as reflecting both different subjects and topics as well as the intensity and depth with which the same topics are taught. These instruction differences originate from tracks that are designed to prepare children for academic higher education at a college or similar institution compared to tracks that prepare children for a more vocational career. Thus, while in principle a large number of school tracks is conceivable, we restrict a tracking

[5]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 a potentially different production technology of cognitive and non-cognitive skills as in Cunha, Heckman, and Schennach (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. In Section 4.1, we use achievement test scores to measure the evolution of this skill measure. As argued in Borghans et al. (2008), achievement test scores measure both cognitive and non-cognitive skills.

system to two school tracks as this corresponds to a typical number across OECD countries.[6] 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. Thus, once a tracking system is established, all children are allocated into one school track and they remain in (a school in) that track for the rest of their schooling years.[7]

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.[8] The pace of instruction in each track is chosen freely by the policymaker and can be set in order to facilitate her goals. For our analysis, we assume that the goal of the policymaker is to maximize aggregate end-of-school skills.[9]

We further assume that there exists a continuum of identical classes (and hence schools) in each track. Thus, if a child is allocated to a particular school track, we can think of her as attending a "representative" class for that track. This implies that all children in a given track are exposed to the same set of classroom and school peers. Again, heterogeneity in instruction paces across tracks do not ex-ante imply different average peer quality across tracks. Instead, this can be the endogenous outcome after sorting.

The school track can then affect next period's skills in two ways: First, through

[6] A detailed comparison between the teaching intensity and learning goals across Germany is provided in Dustmann, Puhani, and Schönberg (2017). We abstract from different probabilities of getting college education after attending a specific school track in this section. This assumption will be relaxed in the quantitative framework in Section 3.

[7] 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, 2013a). While in principle, switches between track during the secondary school years are possible, they are very rare. Dustmann, Puhani, and Schönberg (2017) document a share around 2% of track switches during secondary school in Germany. Track switches are much more common at the end of secondary school, which we allow in the full model in Section 3.

[8] 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.4 , 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.

[9] 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.

interactions with peers within a school track that directly affect a child's skill development. This effect is captured through the average skill level of other children in school track S^c, $\bar{\theta}_{-i|S^c,j^c}$, similar to Duflo, Dupas, and Kremer (2011).[10] Second, school affects child skill development through the pace of instruction in her school track, denoted by P_{j^c,S^c}. In particular, we assume that for each skill level θ_{i,j^c}, there exists a unique optimal pace $P^*(\theta_{i,j^c})$ that maximizes a child i's future skill level, keeping everything else fixed. The optimal pace is assumed to be strictly increasing in current skill, such that higher-skilled children also prefer a higher pace of instruction. Finally, learning decreases monotonically with the distance between a child's optimal teaching pace and the one that she is currently taught at. In combination, these assumptions imply that for a child with a very low current skill level attending a school track 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 track with a very low pace that she actually loses skills.

To illustrate the implications of school tracking policies under these assumptions in an analytically tractable framework, we specify the following form for the child skill technology, where θ_i now refers to the *logarithm* of child i's skill level:[11]

$$\theta_{i,j^c+1} = \theta_{i,j^c} + \alpha\bar{\theta}_{-i|S^c,j^c} + \beta P_{j^c,S^c} - \frac{\delta}{2}P^2_{j^c,S^c} + \gamma\theta_{i,j^c}P_{j^c,S^c} + \eta_{i,j^c+1}, \tag{1}$$

where η_{i,j^c} describe unobserved shocks to the formation of skills. We index the technology by j^c to highlight that the production of skills may be heterogeneous across stages.

In this way, we abstract from any effects of the home environment and parental background on child skill development.[12] We also concentrate on the case with a linear-only direct peer externality, governed by α.[13] Moreover, we assume that the

[10] We abstract from peer effects that operate through friends and the network of a child outside of schools (see Agostinelli, Doepke, et al. (2023)), as our data does not contain information on friendships.

[11] Assuming such a trans-log parametric specification is common in the child skill formation literature (see e.g. Agostinelli and Wiswall (2016). It is convenient as the data measures used to estimate the technology are also assumed to be in log-form. Thus, throughout the paper, we will often refer to "skills" as what is actually the logarithm of skills.

[12] In the quantitative model and the estimation, we account for (a part of) these effects by including parental education directly in the child skill formation technology. Since, we take these as endogenously given, they do not qualitatively change the effects of school policies that we are interested in. We do not consider parent's endogenous investment choices into their child skills.

[13] 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. We note, however that the existence of non-linear peer

distribution of pre-school skills is centered around zero and normal, i.e. $\theta_1 \sim \mathcal{N}(0, \sigma^2_{\theta_1})$. Similarly, the skill shocks η_{i,j^c+1} are assumed to be independently normally distributed around mean zero with a variance $\sigma^2_{\eta_{j^c+1}}$. Finally, we restrict our analysis to cases with $\beta, \gamma, \delta > 0$.

2.2. Full Comprehensive versus Full Tracking System

We start by considering the case with one period of schooling only, so that θ_2 are the skills at the end of school. This means that a tracking system can only be implemented right at the beginning of school by sorting children into two school tracks, in which they stay for the entirety of the school years. Our goal is then to compare the distribution of end-of-school skills in such a full tracking system to that in a full comprehensive system.

To that end, we first describe how policymakers choose the pace of instruction that is taught in every school track. Recall that the policymaker seeks to maximize aggregate end-of-school (log) skills and can set exactly one instruction pace per school track.

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

$$P^*_{1,S^c} = \frac{\beta + \gamma \bar{\theta}_{1,S^c}}{\delta} \tag{2}$$

Proof. Follows from taking the first order condition of the conditional expected value $\mathbb{E}(\theta_2|S^c)$ in (1) with respect to P_{1,S^c} under the i.i.d. assumption of η_2. □

Lemma 1 establishes a necessary condition to maximize aggregate end-of-school skills. It says that the optimal pace a policymaker would pick in every track is given by a function of the average skill level of children in that track, $\bar{\theta}_{1,S^c}$. That is, the policymaker commits to a rule that adjusts the pace of instruction in each track in way, such that it is 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 compared to a world in which she would be taught at her individually optimal level.[14] The average skill level in a track is not known ex-ante but depends on how children are allocated across tracks.

We consider two alternative allocation mechanisms. In the first one, a policymaker can freely allocate all children across tracks. The second alternative consists of

effects could have important implications for the assessment of tracking policies.

[14] 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.

parents making the track decision for the children, with the goal to maximize their end-of-school skill level.[15] That is, children are sorted endogenously into the two tracks, unconstrained by their pre-school skill level. We assume that parents know the skill formation technology (1) as well as the distribution of skill levels among all children and the skill shock distribution. Moreover, it is common knowledge that policymakers commit to setting the pace of instruction according to the rule in Lemma 1.

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 mean.[16] Intuitively, this generates the highest aggregate end-of-school skills as it minimizes the variance of skills in each track. That is, optimal track allocation occurs when the peer groups in each school track are as homogeneous as possible in terms of their skills. 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. The equilibrium of this implied game still features a sharp skill threshold at which expected end-of-school skills are equal in both tracks. The location of this threshold is smaller than the optimal threshold a policymaker would pick, whenever the direct peer effects are positive. Intuitively, parents do not internalize the negative effect that this deviation from the optimal threshold has on aggregate end-of-school skills. The threshold is then characterized by the skill level at which a child is exactly equally well off in both tracks.

Proposition 1. *The allocation of children across tracks is characterized by 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 $\tilde{\theta}_1^* = \mathbb{E}[\theta_1] = 0$.*

[15]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.

[16]A similar argument has been made repeatedly in the theoretical literature, see for instance Epple and Romano (2011).

- *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 $\tilde{\theta}_1^*$.*[17]

Proof. In Appendix. □

Proposition 2 describes the end-of-school distribution of skills in both schooling systems in the one-period model. Provided that $\gamma \neq 0$ and $\delta > 0$, average end-of-school skills in a full optimal tracking system are always larger than in a comprehensive system. While this result depends on the there being only linear direct peer externalities, the exact tracking threshold does not matter. 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. This mirrors the learning-efficiency gain argument that proponents of school tracking typically produce.

The advantage of tracking in terms of increasing aggregate end-of-school skills increases in the complementarity between own skills and instruction pace, γ. The stronger the complementarity the more it pays to stratify children by their skills. Moreover, the advantage increases in the variance of initial child skills $\sigma^2_{\theta_1}$, but decreases in δ, which ultimately governs the concavity of learning with respect to the instruction pace. A full tracking system may, however, lead to larger inequality in end-of-school skills. In particular, condition (3) states that the variance of end-of-school skill might still be larger in a tracking system with positive peer externalities, if tracking occurs at the optimal skill threshold. This is more likely to hold the larger the direct peer externality and the larger the ratio $\frac{\beta\gamma}{\delta}$.

Similarly, 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.[18] 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

[17]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.

[18]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.

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.* $\tilde{\theta}_1$

- *The end-of-school skill distribution in a full tracking system has a "fatter" right tail. In case of tracking at the optimal skill threshold* $\tilde{\theta}_1 = \mathbb{E}(\theta_1)$*, the variance of end-of-school skills in a full tracking system is larger than the variance in a full comprehensive system iff*

$$\alpha^2 + 2\alpha\left(1 + \frac{\beta\gamma}{\delta}\right) - (8-\pi)\frac{\gamma^4}{\pi\delta^2}\sigma^2_{\theta_1} > 0. \tag{3}$$

- *Children with initial skills inside an non-empty interval lose from a full tracking system in terms of their end-of-school skills relative to a full 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. □

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 realize at the end of the period. Therefore, the realization of shocks affects the end-of-school distribution only in that it raises the variance uniformly in both tracks.

2.3. Early versus Late Tracking System

The role of uncertainty changes if we consider a more realistic, 2-period version of (1). That is, we assume that the schooling years consist of two periods, where θ_2 denote the skills at an intermediate stage, and θ_3 are the end-of-school skills. Given initial skills θ_1, the (log) skill formation technology is the one in (1) for $j^c = 1, 2$. At the end of both stages, child skills are subject to an independent skill shock η_{j^c+1}. We maintain the assumption that the policymaker commits to setting the instruction pace in every school track and every period according to Lemma 1. That is, the policymaker has the opportunity to adjust the instruction pace after observing the shock realization η_2.

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. The early tracking system is characterized by an initial track allocation into two tracks, V and A, at the beginning of the school years, $j^c = 1$. As argued before, we do not allow for track switches during the schooling years, including at the beginning of $j^c = 2$. In an early tracking system, a child therefore remains in her school track for the entire school years.[19]

The late tracking system is characterized by all children going to schools that belong to a comprehensive track for the first stage, followed by the allocation into schools that belong to one of two tracks, V and A, at the beginning of the second stage j^c. Importantly, this allocation occurs after the skill shock η_2 is realized.[20]

Proposition 3 shows that average and therefore aggregate end-of-school skills in an optimal late tracking system can actually be larger than in an optimal early tracking system if the variance of the skill shocks are large enough and in particular larger than the variance of initial skills. Intuitively, this represents the key disadvantage of early tracking: since the first track selection cannot be corrected for unexpected skill shocks, the peer groups in each track become more heterogeneous. As a consequence, the average skill levels and hence the instruction paces across tracks are closer together than what would be optimal if re-tracking was possible. Thus, while early tracking produces learning gains in the first stage, it takes away the flexibility to react to unexpected changes in the composition of the tracks. A policymaker seeking to maximize aggregate end-of-school skills may therefore prefer to delay the tracking decision.

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*

$$\frac{\sigma^2_{\eta_2}}{\sigma^2_{\theta_1}} > 1 + \alpha + \alpha^2 + \frac{\beta\gamma}{\delta} + \frac{\beta^2\gamma^2}{2\delta^2} + 2\alpha(1 + \frac{\beta\gamma}{\delta}) + \frac{\gamma^4}{2\delta^2\pi}\sigma^2_{\theta_1}. \tag{4}$$

Proof. In Appendix. □

As evident from (4), the quantitative significance of this motive depends not only on estimates of all child skill formation technology parameters, which are likely a

[19] 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 use only relatively rarely.

[20] We do not consider a full 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.

function of the age of the child and the school stage, but also crucially on the size of the skill uncertainty. Moreover, so far, we have only been concerned with the implications of the child skill technology in different school tracking scenarios on the distribution of end-of-school skills.[21] The goal of this paper, however, is to provide insights on effects of such policies on life-time and intergenerational outcomes. To that end, there are two further important aspects about school tracking that warrant consideration.

The first concerns the track choice made by parents. Research on school tracking has found that parents with higher socio-economic status (SES) are more likely to send their child to an academic track school than parents with a lower socio-economic status, even conditional on school performance or achievement test scores prior to the track decision (Falk, Kosse, and Pinger, 2021). For example, in our data on Germany, we find that this conditional gap in academic track attendance is around 24 percentage points. Moreover, around 20% of parents who themselves have college education overrule a recommendation of primary school teachers recommending their child to be sent to a vocational track school.[22]

There may be multiple reasons behind this own-track bias of parents. For example, parents may be able to better support their child in a track that they are more familiar. Parents may also just over-, or underestimate the potential of their children or have preferences for their child following in their own footsteps.[23] Finally, parents likely make their track decision considering not just end-of-school skills but also future labor market prospects. Given that graduation from an academic track school is often a catalyst to tertiary education at a university or college, the track decision is thus consequential for the potential wage that can be earned with an academic degree. However, this wage is a function of labor demand by firms who likely employ labor from both vocational and academic degrees. Thus, in a general equilibrium setting, parents need to take into account that, if everyone sends their child to the academic track, not only with aggregate learning suffer, but also wages in academic jobs will be suppressed.

[21]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.

[22]See Appendix A.2 for some reduced-form evidence on the school track selection by parental background, deviations from track recommendations and the consequences of such deviations in terms of later learning outcomes.

[23]For example, if the goal is to take over a family firm. Of course, it is also conceivable that parents just have better knowledge of their child's underlying potential which prompts them to deviate from the recommended track.

Whatever the exact reason for deviation of parent's track choice from the recommended path, it is clear that such biases in track choice may lead to misallocation of children across tracks. For example, a child with low underlying true potential could be sent to the academic track by parents that have reasons or preferences to prefer this track. This would lead to learning losses not only for the individual child but also pose an externality for all other children as the instruction pace is endogenous to the peer composition.

The second aspect concerns the possibility of second-chance opportunities during the school career. As described earlier, track switches during secondary school education are very rare. However, while typically only graduation from an academic track school guarantees the possibility of access to university or college higher education, in reality many pathways through the education system exists.[24] While the first track choice remains an important predictor for example for college attendance, the existence of such second-chance opportunities likely affects the quantitative importance of skill shocks as they provide a chance to "correct" earlier mistakes.

We address these two issues in our full quantitative model that we describe next.

3. Quantitative General Equilibrium 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.[25] This frequency allows us to investigate meaningful variations in school tracking ages. The dynastic structure implies that there are 19 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.[26]

In period $j^c = 1$ a 2-year old child enters into a one-parent household, equipped with an initial learning ability ϕ^c, which is imperfectly transmitted from her parent Yum (2022). This learning ability translates into a pre-school child skill level θ_2. For $j^c = 1, ..., 4$, the parent makes all decisions for the household. These include

[24] See Dustmann, Puhani, and Schönberg (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 beginning of possible tertiary education that is characterized by multiple possibilities. In fact, in their paper these second-chance opportunities are the main reason why the initial track choice does not have an impact on the marginal child.

[25] 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.

[26] We denote by the superscript c that a variable is associated with the child generation in a household, starting from her birth until she leaves the household. Variables without superscript are associated with the decision-maker in all stages.

the decision in which school track to enroll the child. In line with the arguments in Section 2, we assume that there are two representative school tracks in the model economy, a vocational school track, $S^c = V$ and an academic school track $S^c = A$. Once tracked, a child finishes secondary school in the same track that she started in. In the baseline model, the tracking decision happens at the beginning of secondary school in $j^c = 3$, when children are 10 years old.

The development of child skills during the schooling years is governed by the technology in (1). We keep assuming that classrooms and schools across tracks are identical and that policymakers set the instruction pace in each track optimally according to Lemma 1. Then, following the arguments in Section 2, we can write next period's child skills in general as

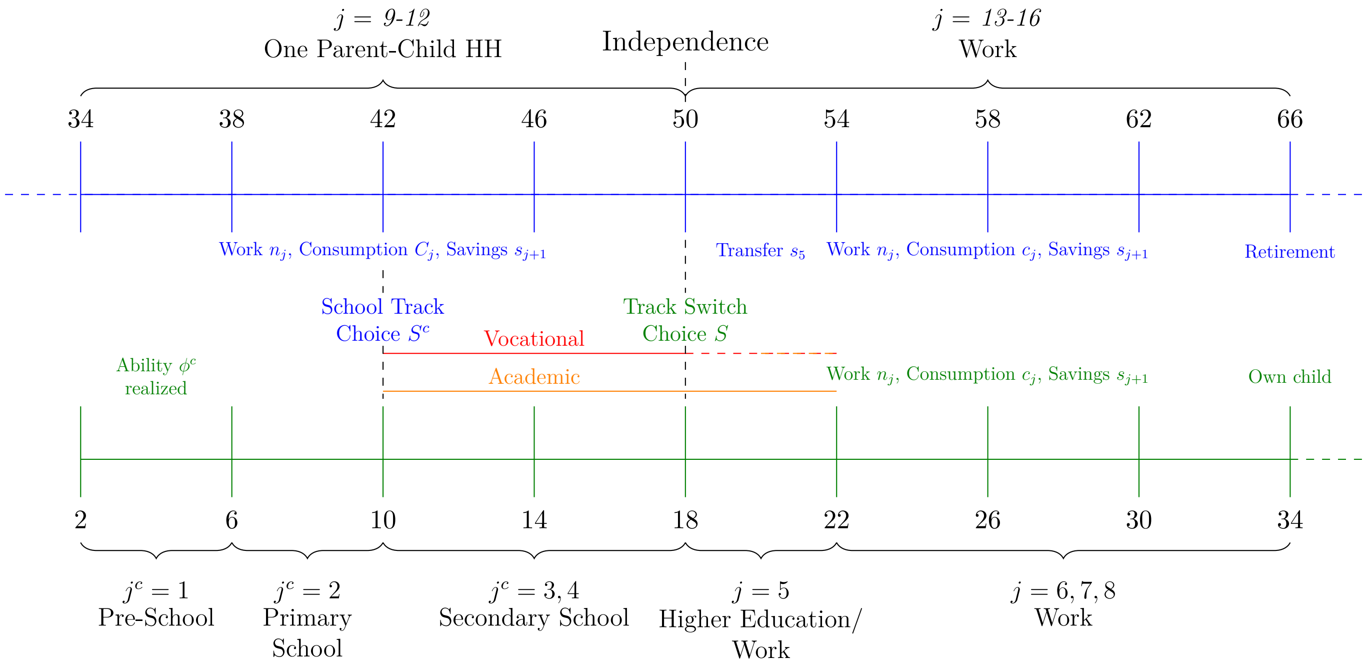
$$\theta_{j^c+1} = g_{j^c}(\theta_{j^c}, S, \bar{\theta}_{S^c,j^c}, \bar{\theta}^2_{S^c,j^c}, \eta_{j^c+1}) \tag{5}$$

That is, we can replace the dependence of child skills on the instruction pace with the average peer skills in school track S^c and its squared term. Moreover, we allow for the explicit dependence of child skill development on parental education S. This allows us to capture effects from the household environment but also parental inputs such as monetary investments into the child skill development.

In period $j = 5$, the child finishes school and becomes an independent adult. She can decide whether to switch tracks and pursue academic higher (college) education or obtain a vocational/professional degree. This allows for the possibility of "second chances" for high-skilled children to obtain an academic degree while at the same time accounting for the fact that not all graduates from academic school tracks go on to study in college. The higher education choice is consequential in that it qualifies an individual to supply either vocational or academic labor to a representative firm, which uses both types of labor as inputs. The initial level of adult human capital of each individual then depends on her end-of-school skills, her higher education and her innate ability. Obtaining academic higher education constrains the time available to work – providing still vocational level labor – as it generally takes longer than obtaining a vocational degree.

The periods $j = 6, 7, 8$ are the working years of an agent whose own child is not born yet. The periods $j = 9, ..., 16$ encompass the parenting years of each agent as well as the working years after the own child left the household. During these periods, she makes consumption, savings and labor supply decisions and 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 $j = 17, ..., 19$, the individual is retired and makes consumption and savings

Figure 1: Timeline of Life-cycle Events



decisions while earning retirement benefits. Everyone dies with certainty after model period $j = 19$, that is at age 82.

Throughout their lives, there are a number of frictions that can affect the decisions of the model agents. First, markets are incomplete as both adult human capital as well as child skills are subject to uninsurable idiosyncratic uncertainty. Second, in each period, individuals can only borrow an amount that is constrained by what they can fully repay in the next period using a government lump-sum transfer (Lee and Seshadri, 2019). Moreover, when deciding on inter-vivos transfers, parents cannot borrow against their child's future income. Third, in line with the discussion in Section 2, parents bias the choice of which track to send their child to towards the same school track that they themselves went to. We can view this as the result from an informational friction. Finally, the possibility to obtain college education is constrained by a child's school track and her end-of-school skills (and luck). This reflects the common practice that universities require minimum entry requirements.[27]

In the following, we present the recursive formulation of the agents decisions in more detail.

[27] In principle, only graduating from the academic school track allows for direct access to college education. However, in practice, there are a number of second chance options through which individuals with a vocational school track degree can gain access to academic higher education institutions.

3.1. Individual Problems

Period $j = 5$: Independence and Higher Education

At the beginning of the period, the state space of the newly independent adult comprises the secondary school track she graduated from, $S^c \in \{V, A\}$, where V refers to the vocational school track and A denotes the academic school track, her end-of-school skills θ_5, and an initial savings level s_5 which she obtains as a financial transfer from her parent and takes as given. In addition, the state space contains the higher education state of her parent, which we denote by $S^p \in \{V, A\}$.[28] The last state, ϕ, denotes innate ability of the individual which will be transferred (imperfectly) to her child. End-of-school skills, re-scaled with a constant factor ξ, determine deterministically the first adult human capital level, h_5 which is rewarded in the labor market.

The first decision an independent adult makes after leaving secondary school, is then whether to obtain academic higher (college) education $S = A$ or start working in a vocational profession $S = V$. Obtaining college education gives an individual access to the labor market for college educated workers, which pays an effective wage rate w_A and results in an exogenous growth rate of human capital γ_A that may be different from the one achieved during a vocational career, γ_A. During college education, an individual may also work at the vocational wage rate w_V. However, obtaining a college education reduces the time available for work in $j = 5$, as individuals spend part of their total time endowment of 1 studying, $\bar{n}(S = A) < 1$.

On top of that, going to college occurs a non-pecuniary cost $\psi(S^c, \theta_5, \nu(S^p))$. This cost is dependent on the secondary school track S^c, the end-of-school skill level θ_5, as well as the realization of an idiosyncratic "college taste" shock.[29] The purpose of this cost is to account for the fact that it is in principle possible to obtain college education even after not graduating from an academic track secondary school. However, college education through such "second-chance" opportunities is much less frequent.[30] Moreover, it is a salient feature in the data that students with higher end-of-school skills are more likely to go to (and graduate from) college, regardless of their secondary school track. For that reason, we allow the college cost ψ to be

[28] We add the superscript p here as the child is already independent and decides on her own education S, which may be affected by her parent's education.

[29] These type of "psychic costs of higher education" are common in the literature (see e.g. Abbott et al. (2019) and Daruich (2022)).

[30] 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.

dependent on the end-of-school skill level. Finally, the taste shock $\nu(S^p)$ may 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. 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.

Then, conditional on the realization of the stochastic college taste shock $\nu(S^p)$ the value of an agent at the beginning of $j = 5$ is defined by

$$\begin{aligned} V_5(S^c, \theta_5, s_5, \phi^c, \nu(S^p)) = \max\{&W_5(S = V, h_5, s_5, \phi), \\ &W_5(S = A, h_5, s_5, \phi) - \psi_A(S^c, \theta_5, \nu(S^p))\} \\ \nu \sim \mathcal{N}(&\mu_{\nu,S^p}, \sigma_\nu^2), \end{aligned} \tag{6}$$

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

$$\begin{aligned} W_5(S, h_5, s_5, \phi) = &\max_{\substack{c_5>0, s_6 \geq \underline{s}, \\ n_5 \in [0, \bar{n}(S)]}} \left\{ u(c_5, n_5) + \beta \int V_6(S, h_6, s_6, \phi) dF(\varepsilon_6) \right\} \\ \text{s.t. } c_5 + s_6 &= y_5 + (1 + r)s_5 - T(y_5, s_5) \\ y_5 &= w_V h_5 n_5 \\ h_5 &= \xi \exp(\theta_5) \\ h_6 &= \gamma_{5,S} h_5 \varepsilon_6 \\ \phi &= \phi^c \\ \log \varepsilon_j &\sim \mathcal{N}(0, \sigma_\varepsilon^2). \end{aligned} \tag{7}$$

Thus, the individual chooses optimal consumption c_5 and savings s_6, subject to the life-cycle borrowing constraint $\underline{s}$, as well as hours worked n_5. We denote by y_5 the gross labor income, which is given by $w_V h_t n_5$. We assume that all work during the higher education stage is of the vocational nature and therefore earns a wage per efficient labor unit of w_V. The function $T(y_5, s_5)$ gives taxes net of transfers which consist of labor income and capital taxes. The interest rate on savings is r. All prices implicitly depend on the distribution of agents in the economy, which we suppress for notational convenience. Initial human capital h_5 corresponds to end-of-school skills, shifted by the constant ξ. Future human capital grows at the education-specific rate γ and 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, Ventura, and Yaron (2011). All future values are discounted by β.

Periods $j = 6, 7, 8$: Young adults without child

In this life stage, the model agents have completed all secondary and higher education but do not have own children yet. They choose optimal consumption, savings, and labor supply to solve the standard life-cycle problem. For $j = 6, 7$ the values are hence given by

$$\begin{aligned} V_j(S, h_j, s_j, \phi) = & \max_{\substack{c_j>0, s_{j+1}\geq \underline{s}, \\ n_j\in[0,1]}} \left\{ u(c_j, n_j) + \beta \int V_{j+1}(S, h_{j+1}, s_{j+1}, \phi) dF(\varepsilon_{j+1}) \right\} \\ \text{s.t. } c_j + s_j &= y_j + (1+r)s_j - T(y_j, s_j) \\ y_j &= w_S h_j n_j \\ h_{j+1} &= \gamma_{j,S} h_j \varepsilon_{j+1}. \end{aligned} \tag{8}$$

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

$$\begin{aligned} V_8(S, h_8, s_8, \phi) = & \max_{\substack{c_8>0, s_9\geq \underline{s}, \\ n_9\in[0,1]}} \left\{ u(c_8, n_8) + \beta \int V_9(S, h_9, s_9, \phi^c) dF(\varepsilon_9) dG(\phi^c|\phi) \right\} \\ \text{s.t. } c_8 + s_9 &= y_8 + (1+r)s_8 - T(y_8, s_8) \\ y_8 &= w_S h_8 n_8 \\ h_9 &= \gamma_{9,S} h_8 \varepsilon_9 \\ \log \phi^c &= \rho_\phi \log \phi + \epsilon_\phi, \qquad \epsilon_\phi \sim \mathcal{N}(0, \sigma_\phi^2). \end{aligned} \tag{9}$$

Period $j = 9$: Parent with newborn child

As the newborn child enters, a household is now comprised of a parent and a child. We denote by q an adult consumption-equivalent scale so that c_9 remains the consumption of the household (Yum, 2022). The child's innate learning ability ϕ^c is realized and becomes the relevant state variable. It determines next period's log child skills θ_2, which are the skills just before entering primary school. In the data, these child skills are correlated with the parental background, which likely reflects both genetic components and early-childhood household effects. In our model, the correlation of ϕ between parents and children thus captures both these channels.

$$
\begin{aligned}
V_9(S, h_9, s_9, \phi^c) = & \max_{\substack{c_9>0, s_{10}\geq \underline{s}, \\ n_9\in[0,1]}} \left\{ u(\frac{c_9}{q}, n_9) + \beta \int V_{10}(S, h_{10}, s_{10}; \theta_2, \phi^c) dF(\varepsilon_{10}) \right\} \\
\text{s.t. } c_9 + s_{10} &= y_9 + (1+r)s_9 - T(y_9, s_9) \\
y_9 &= w_S h_9 n_9 \\
h_{10} &= \gamma_{9,S} h_9 \varepsilon_{10} \\
\theta_2 &= \log \phi^c, \qquad \eta_2 \sim \mathcal{N}(\mu_{2,S}, \sigma^2_{\eta,2})
\end{aligned}
$$

Period $j = 10$: Parent with child in primary school

In this period, every child goes to comprehensive primary school. Throughout all school years, the evolution of log child skills are then given by the learning function $g_{j^c}(\theta_{j^c}, S^p, \bar{\theta}_{S^c,j^c}, \theta^2_{j^c}, \bar{\theta}^2_{S^c,j^c}, \eta_{j^c+1})$ that we estimate directly from the data. That is, under the assumption of optimal instruction pace setting, future log child skills now depend on a quadratic in current skills and the squared distance between own skills and average skill level of all children that go to a school that belongs to the same track. In addition, child skill development also directly depends on the parental education level S as a proxy for parental inputs and direct effects of different home environments. Finally, the skills shocks η_{j^c+1} are drawn from the distribution $H(\eta_{j^c+1})$, which we assume is a normal distribution in logs, with zero mean and period-specific variance $\sigma^2_{\eta_{j^c+1}}$.

Since the average skill level in a school track depends on the distribution of children across tracks, parents need to form expectations over these, which in equilibrium, must coincide with the actual distributions. As with prices r, and w_S, we keep the dependence of average skill levels on the aggregate distribution implicit. During primary school, there is only one comprehensive school track, such that the log child skill evolution depends on the overall average skill level among all children in that period. The value at this stage is then

$$V_{10}(S,h_{10},s_{10};\theta_2,\phi^c)=\max_{\substack{c_{10}>0,s_{11}\geq \underline{s},\\ n_{10}\in[0,1]}}\left\{\begin{array}{c} u(\frac{c_{10}}{q},n_{10})+ \\ \beta\int V_{11}(S,h_{11},s_{11};\theta_3,\phi^c)dF(\varepsilon_{11})dH(\eta_3)\end{array}\right\}$$

$$\begin{aligned}
\text{s.t. } c_{10}+s_{11} &= y_{10}+(1+r)s_{10}-T(y_{10},s_{10}) \\
y_{10} &= w_S h_{10} n_{10} \\
h_{11} &= \gamma_{10,S} h_{10}\varepsilon_{11} \\
\theta_3 &= \kappa_{0,2}+\kappa_{1,2}\theta_2+\kappa_{2,2}\theta_2^2+\kappa_{3,2}\bar{\theta}_2+\kappa_{4,2}(\theta_2-\bar{\theta}_2)^2+\kappa_{5,2}S+\eta_3. \\
\eta_3 &\sim \mathcal{N}(0,\sigma^2_{\eta,3}).
\end{aligned} \tag{10}$$

Period $j = 11$: Parent makes school track decision

In the beginning of the period, 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, $S^c \in \{V, A\}$. This decision is affected by the value of placing her child in each track, W_{11}, but also by a fixed preference shifter, $\chi(S, S^c)$, that depends on the child's but also on the parent's track. In the data, the school track selection by parents is significantly tilted towards sending your child to the same track that the parent went to, even conditional on child skills prior to the track decision. In particular, parents systematically deviate from the recommended track for their child, where the recommendation typically comes from the primary school teacher. We do not explicitly micro-found these deviations. Instead we calibrate the preference shifters, such that the model replicates the deviations from teacher recommendation, which we assume are the optimal track choices in the absence of $\chi(S, S^c)$. Then the value at the beginning of period 11 is given by

$$V_{11}(S,h_{11},s_{11};\theta_3,\phi^c)=\max_{S^c\in\{V,A\}}\left\{W_{11}(S,h_{11},s_{11};S^c,\theta_3,\phi^c)-\chi(S,S^c)\right\}, \tag{11}$$

where the values of sending your child to a school that belongs to school track S^c

are given by

$$W_{11}(S,h_{11},s_{11};S^c,\theta_3,\phi^c) = \max_{\substack{c_{11}>0, s_{12}\geq \underline{s}, \\ n_{11}\in[0,1]}} \begin{Bmatrix} u(\frac{c_{11}}{q}, n_{11}) + \\ \beta \int V_{12}(S,h_{12},s_{12};S^c,\theta_4,\phi^c) dF(\varepsilon_{12}) dH(\eta_4) \end{Bmatrix}$$

$$\begin{aligned} \text{s.t. } c_{11} + s_{12} &= y_{11} + (1+r)s_{11} - T(y_{11}, s_{11}) \\ y_{11} &= w_S h_{11} n_{11} \\ h_{12} &= \gamma_{11,S} h_{11} \varepsilon_{12} \\ \theta_4 &= \kappa_{0,3} + \kappa_{1,3}\theta_3 + \kappa_{2,3}\theta_3^2 + \kappa_{3,3}\bar{\theta}_{3,S^c} + \kappa_{4,3}(\theta_3 - \bar{\theta}_{3,S^c})^2 + \kappa_{5,3}S + \eta_4. \end{aligned} \tag{12}$$

Note that future child skills are now affected by $\bar{\theta}_{3,S^c}$, which are the average skill levels among children in school track S^c.

Period $j = 12$: Parent with child in secondary school

The child remains in the school track chosen in the prior period. That is, we assume that there is no track switching possibility during secondary school.[31]

$$V_{12}(S,h_{12},s_{12};S^c,\theta_4,\phi^c) = \max_{\substack{c_{12}>0, s_{13}\geq \underline{s}, \\ n_{12}\in[0,1]}} \begin{Bmatrix} u(\frac{c_{12}}{q}, n_{12}) + \\ \beta \int V_{13}(S,h_{13},s_{13};S^c,\theta_5,\phi^c) dF(\varepsilon_{13}) dH(\eta_5) \end{Bmatrix}$$

$$\begin{aligned} \text{s.t. } c_{12} + s_{12} &= y_{12} + (1+r)s_{12} - T(y_{12}, s_{12}) \\ y_{12} &= w_S h_{12} n_{12} \\ h_{13} &= \gamma_{12,S_p} h_{12} \varepsilon_{13} \\ \theta_5 &= \kappa_{0.4} + \kappa_{1,4}\theta_4 + \kappa_{2,4}\theta_4^2 + \kappa_{3,4}\bar{\theta}_{4,S^c} + \kappa_{4,4}(\theta_3 - \bar{\theta}_{4,S^c})^2 + \kappa_{5,4}S + \eta_5. \end{aligned} \tag{13}$$

Period $j = 13$: Parent when child becomes independent

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, s_5, while taking into account the child's future value $V_5(S^c, \theta_5, s_5, \phi^c, \nu(S))$. As in Daruich (2022), we model this as an interim decision problem and assume that the parent already knows

[31] 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 track. The great majority of track switches are from an academic track school to a vocational track school rather than the other way.

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. We assume that these taste shocks follow a normal distribution around zero, with variance σ_ν^2. As is common, the transfer cannot be negative, so that parents cannot borrow against the future income of their child. The strength of the parental altruism motive is governed by δ, which weights how much parents take the continuation value of their child into account. The value at the beginning of period 13 is then

$$\begin{aligned} V_{13}(S, h_{13}, s_{13}; S^c, \theta_5, \phi^c) &= \max_{s_5 \geq 0} \left\{ \widetilde{V}_{13}(S, h_{13}, s_{13} - s_5) + \delta \, \mathbb{E}_\nu \, V_5(S^c, s_5, \theta_5, \phi^c, \nu(S)) \right\} \\ \nu &\sim \mathcal{N}(\mu_{\nu,S}, \sigma_\nu^2), \end{aligned} \tag{14}$$

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

$$\begin{aligned} \widetilde{V}_{13}(S, h_{13}, s_{13}) &= \max_{\substack{c_{13} > 0, s_{14} \geq \underline{s}, \\ n_{13} \in [0,1]}} \left\{ u(c_{13}, n_{13}) + \beta \int V_{14}(S, h_{14}, s_{14}) dF(\varepsilon_{14}) \right\} \\ \text{s.t. } c_{13} + s_{14} + s_5 &= y_{13} + (1+r)s_{13} - T(y_{13}, s_{13}) \\ y_{13} &= w_S h_{13} n_{13} \\ h_{14} &= \gamma_{13,S} h_{13} \varepsilon_{14}. \end{aligned} \tag{15}$$

Periods $j = 14, 15, 16$ Remaining work life and retirement

During these periods the agent continues to work, where the value functions are given by

$$\begin{aligned} V_j(S, h_j, s_j) &= \max_{\substack{c_j > 0, s_{j+1} \geq \underline{s}, \\ n_j \in [0,1]}} \left\{ u(c_j, n_j) + \beta \int V_{j+1}(S, h_{j+1}, s_{j+1}) dF(\varepsilon_{j+1}) \right\} \\ \text{s.t. } c_j + s_{j+1} &= y_j + (1+r)s_j - T(y_j, s_j) \\ y_j &= w_S h_j n_j \\ h_{j+1} &= \gamma_{j,S} h_j \varepsilon_{j+1} \quad \text{if } j < 16 \\ h_{j+1} &= h_j \varepsilon_{j+1} \quad \text{if } j = 16. \end{aligned} \tag{16}$$

Note that in the period prior to retirement, $j = 16$, the agent need not take expectations over future market luck shocks any more.

Periods $j = 17, 18, 19$ Retirement

Everybody retires at the beginning of model period 17, which corresponds to age 66 in real life. In the remainder of her life, the agent receives retirement benefits $\pi_j(h_j)$, which depend on the last human capital level before retirement.[32] After model period 19, that is at age 82 agents die and exit the model. The values in these periods are

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

Finally, the value of death is normalized to zero.

3.2. Aggregate Production, Government, and Equilibrium

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:

$$H = [\omega H_V^\epsilon + (1 - \omega)H_A^\epsilon]^{\frac{1}{\epsilon}}. \tag{18}$$

Here, H_V is the aggregate labor supply in efficiency units of workers with vocational higher education, and H_A that of workers with academic (college) higher 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, Storesletten, and Violante, 2017). It also taxes capital income linearly according to $\tau_s r s_j$ (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.

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 19 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_V and w_A

[32] 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.

are competitively determined and move to clear all markets.

A special feature of our model is that learning during the school years depends on the distribution of children across school tracks. Thus, an equilibrium in particular requires that parents perceptions about the average skill level in each track when making the track choice corresponds to the actual one. 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 detailed definition of the equilibrium is given in Appendix A.3.

4. Estimation and Parameterization

The model is designed to fit the German Education System, an overview of which is given in Appendix A.4. 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 3 and of the internally estimated ones in Table 4.

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), which comprises detailed longitudinal data on the educational process, acquired competences, as well as the learning environment and context persons of six cohorts of children in nationally representative samples in Germany (Blossfeld, Roßbach, and Maurice, 2011). We select the samples of starting cohorts 2, 3, and 4, which cover children starting in kindergarten (around the age of 4), lower secondary school (around age of 10) and upper secondary school (around age 15) in 2010 and followed annually until 2018. A key component of the information collected are regular standardized assessment tests of the children's competencies in areas such as mathematics, reading, sciences, vocabulary, or grammar.[33] 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.

[33] We provide detailed information on these tests in Appendix A.5.

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.

4.1. Estimation of the the Child Skill Formation Technology

We specify the production technology of (the logarithm) of child i's skills (5) that we take to the data as follows:[34]

$$\begin{aligned}\theta_{i,j^c+1} = \kappa_{0,j^c} + \kappa_{1,j^c}\theta_{i,j^c} + \kappa_{2,j^c}\theta_{i,j^c}^2 + \kappa_{3,j^c}\bar{\theta}_{-i|C_{S^c}(i),j^c} \\ + \kappa_{4,j^c}(\theta_{i,j^c} - \bar{\theta}_{S^c(i),j^c})^2 + \kappa_{5,j^c}S_i + \eta_{i,j^c+1}.\end{aligned} \tag{19}$$

Note that (19) is just a rearranged version of the child skill technology (1) we developed in Section 2 to illustrate the theoretical effects of school tracking on end-of-school skills, after substituting in the optimal pace of instruction in a child's school track and with three modifications. First, we add the education of the parent, S, as a regressor. As discussed earlier, the dependence on S allows us to (partially) capture the possibility that a child's skill development depends directly on their parental background or on parental inputs that are correlated with their socio-economic status. In the estimation, S is a time-constant dummy that equals 1 if child i comes from a household in which at least one parent is college educated.

Second, we denote by $\bar{\theta}_{-i|C_{S^c},j^c}$ the average skill level of child i's *classroom* peers, as opposed to $\bar{\theta}_{S^c(i),j^c}$, which refers to the average skill level of all children in a school

[34]Following the work in Cunha, Heckman, and Schennach (2010), much of the empirical and quantitative literature using child skill formation technologies employed parametric specifications of (??) 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 Daruich (2022) and Fuchs-Schündeln, Krueger, Kurmann, et al. (2021). To retain tractability, we follow Agostinelli and Wiswall (2016) and opt for the trans-log approach.

that belongs to track S^c and arises from optimal track-specific instruction paces. Note that in the model, $\bar{\theta}_{-i|C_{S^c(i)},j^c} = \bar{\theta}_{S^c(i),j^c}$, 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. Moreover, using classroom-specific direct peer effects aids in the identification of (19). This is because, as in the model, we consolidate schools into a maximum of two school tracks in the data, which, as discussed in Section A.4 resembles reality in Germany over the past decade. Thus, a model that includes $\bar{\theta}_{S^c,j^c}$, $\bar{\theta}^2_{S^c,j^c}$ and the interaction $\theta\bar{\theta}_{S^c,j^c}$ as separate regressors is not identified because of collinearity. 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 (19), which is why we drop θ^2.[35]

The third modification is that we follow the child skill formation literature and allow the parameters $\kappa_n, n = 0, ...5$ to be stage-specific as is common in the literature. The parameters of (19) then relate to those in (1) according to $\kappa_0 = \frac{\beta^2}{2\delta}$, $\kappa_1 = (1+\frac{\beta}{\gamma}\delta)$, $\kappa_2 = \alpha$, and $\kappa_3 = -\kappa_4 = \frac{\gamma^2}{2\delta}$.[36] We formally test the hypothesis of $\hat{\kappa}_3 = -\hat{\kappa}_4$ after the estimation.

As is common in the child skill formation literature (Agostinelli and Wiswall, 2016; Cunha, Heckman, and Schennach, 2010), we think of 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 that is given by

$$M_{i,k,j^c} = \mu_{k,j^c} + \lambda_{k,j^c}\theta_{i,j^c} + \epsilon_{i,k,j^c}, \tag{20}$$

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

[35]This is also the reason why we prefer (19) over a model that includes $\bar{\theta}^2_{S^c,j^c}$ and the interaction $\theta\bar{\theta}_{S^c,j^c}$ as separate regressors, such as (1), even when using class-specific peer effects. While in version (19), we just have to drop the squared term on own skills, which is typically statistically insignificant even when two tracks are available, in version (1), we cannot identify neither the coefficient in front of $\bar{\theta}^2_{S^c,j^c}$ nor that of $\theta\bar{\theta}_{S^c,j^c}$.

[36]See equation (A.2) in Appendix A.1 for the rearrangement.

which we keep implicit.

Following Cunha, Heckman, and Schennach (2010), we normalize $\mathbb{E}(\theta_{j^c}) = 0$ and $\lambda_{1,j^c} = 1$ for all j^c. That is the first factor loading is normalized to 1 in all periods.[37] We further normalize the measurement errors, such that $E(\epsilon_{k,j^c}) = 0$ for all j^c. Given that, the location parameters μ_{k,j^c} are identified from the means of the measures. In order to identify the factor loadings, we further assume that the measurement errors are independent from 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 period by ratios of covariances of the measures (as in Agostinelli, Saharkhiz, and Wiswall (2019)):

$$\lambda_{k,j^c} = \frac{Cov(M_{k,j^c}, M_{k',j^c})}{Cov(M_{1,j^c}, M_{k',j^c})} \tag{21}$$

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

$$\theta_{i,j^c} = \frac{M_{i,k,j^c} - \mu_{k,j^c}}{\lambda_{k,j^c}} - \frac{\epsilon_{i,k,j^c}}{\lambda_{k,j^c}} = \widetilde{M}_{i,k,j^c} - \frac{\epsilon_{i,k,j^c}}{\lambda_{k,j^c}}. \tag{22}$$

Equipped with identified latent variables up to measurement error for all periods, we can plug these into the child skill technology (19), which yields

$$\begin{aligned} \widetilde{M}_{i,k,j^c+1} = \kappa_{0,j^c} + \kappa_{1,j^c}\widetilde{M}_{i,k,j^c} + \kappa_{2,j^c}\widetilde{M}^2_{i,k,j^c} + \kappa_{3,j^c}\overline{\widetilde{M}}_{-i|C_{S^c}(i),j^c} \\ + \kappa_{4,j^c}(\widetilde{M}_{i,k,j^c} - \overline{\widetilde{M}}_{S^c(i),j^c})^2 + \kappa_{5,j^c}S_i + \zeta_{i,k,j^c+1}, \end{aligned} \tag{23}$$

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

Importantly, the residual ζ_{i,k,j^c+1} now contains not only structural skill shocks, η_{i,j^c+1}, but also the measurement errors, ϵ_{i,k,j^c} 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

[37] 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.

of (23) will be biased. To account for that, we follow the literature and use excluded measures as instrumental variables, which we describe in Appendix A.6.[38]

We present our preferred IV estimates of the child skill production technology parameters for all three stages of the schooling career in Table 1 and the baseline OLS estimates in Appendix Table A.4. The IV estimates are typically larger than the OLS estimates. This indicates the importance of measurement error and our correction for it. Recall that θ_{i,j^c} 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 Agostinelli, Saharkhiz, and Wiswall (2019) and Cunha, Heckman, and Schennach (2010)).

Note, that this own-skill elasticity ignores the effect coming through θ^2_{i,j^c}. As discussed before, θ^2_{i,j^c} cannot be identified in primary school as there is only one comprehensive track. During secondary school, the estimated coefficient $\hat{\kappa}_2$ are positive and significant at least at the 10% level. More importantly though, we cannot reject the hypothesis that $\hat{\kappa}_2 = -\hat{\kappa}_4$. Thus, as conjectured, the effects of θ^2_{i,j^c} are likely to cancel out.

The estimated $\hat{\kappa}_3$ are generally small and often statistically insignificant. Thus, our empirical estimates do not provide evidence in favor of positive direct peer externalities in the classroom. In Appendix A.7, 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.

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 associate with an up to 0.72% decrease in the next periods skill. 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.

[38] Under the assumption that measurement error is uncorrelated across measures, this strategy will take of measurement error and the interaction terms included in (23) 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, Heckman, and Schennach (2010).

Taken together, the estimated elasticities κ_n, $n = 1, ..., 4$ imply that the parameters in the child skill technology we used to illustrate the theoretical effects of school tracking policies are, with the exception of α all positive, as conjectured.[39] This provides us reassurance that the child skill technology we develop not only entails sensible predictions that reflect existing empirical evidence from different contexts and popular arguments about school tracking policies, but also is based on parameter restrictions that are born out in the data.

The final estimates we use to parameterize the child skill formation technology in our model are the sets κ_{n,j^c} for $n = 0, 1, 2, 4, 5$ and $j^c = 2, 3, 4$ coming from the IV estimates in Table 1. That is we do not include a direct peer externality in the model, given the insignificant estimates of κ_{3,j^c}. For the first stage of secondary school, $j^c = 3$ we opt to use the parameter estimates from the age 12-14/15 sample.

Table 1: IV 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
$\hat{\kappa}_{1,j}$	$\theta_{i,j}$	0.954	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)
$\hat{\kappa}_{3,j}$	$\bar{\theta}_{-i,C_{S(i),j}}$	-0.043	0.283	-0.199	-0.08
		(0.069)	(0.212)	(0.133)	(0.108)
$\hat{\kappa}_{4,j}$	$(\theta_{i,j} - \bar{\theta}_{S(i),j})^2$	-0.215	-0.559	-0.719	-0.596
		(0.136)	(0.359)	(0.325)	(0.204)
$\hat{\kappa}_{5,j}$	$S = A$	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

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

[39] For example, the estimated coefficients using the age 12-14/15 sample in model period $j = 3$, imply parameters of $\beta = 0.178, \gamma = 1.175$, and $\delta = 2.72$.

4.2. Remaining Parameters

Preferences

We specify the following period utility function of the household.

$$u(c_j, n_j) = \frac{c_j^{1-\sigma}}{1-\sigma} - b\frac{n_j^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}}. \tag{24}$$

We set the inverse elasticity of intertemporal substitution, σ to 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 total time available after sleep and self-care is normalized to 1.

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 ranges 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 target a ratio of 0.6.

Finally, as discussed before, the school track decision by parents is characterised by a bias towards the own track in the data. To reflect these in the model, we use the preference shifter $\chi(S, S^c)$ at the time of the track decision. In particular, we assume that $\chi(S, S^c) = \chi_A \mathbb{I}\{S = A \wedge S^c = A\} + \chi_V \mathbb{I}\{S = V \wedge S^c = V\}$, so that the bias of academic parents is governed by χ_A and that of vocational parents by χ_V. We estimate these parameters to match the share of deviations from secondary school track recommendations by parental education in the data. These recommendations are typically given by the primary school teachers before transition to secondary school. They are based on both a reflection of the child's achievement during primary school as well as the teachers assessment of the academic potential and success probability of the child in an academic track school. Thus, we would argue that the recommendations are forward-looking and, since the primary school teachers typically observe the children over multiple years every day during the week, based on a similar information set as the parents possess. For that reason, we think of the recommended school track in the model as the one which a parent would have chosen without the own-track bias. Then, deviations from that unbiased track choice by parental education map into deviations from teacher recommendation.

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 pre-school child skills by parental education, that is $(\bar{\theta}_{2|S=A} - \bar{\theta}_{2|S=V})/Std(\theta_2)$. 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 2, the size of such shocks has important implications about 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 σ^2_{η,j^c+1}, for $j^c = 2, 3, 4$. As target moments, we choose the correlation of a child's math test score percentile rank across periods. 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. In reality, such changes may 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.

College Costs

We specify the following functional form for the college costs $\psi_A(S^c, \theta_5, \nu(S^p))$:

$$\psi_A(S^c, \theta_5, \nu) = \exp(\psi + \psi_{S_c=V} + \psi_\theta \theta_5 + \nu)$$
$$\nu \sim \mathcal{N}(\mu_{\nu,S^p}, \sigma^2_\nu).$$

This formulation represents two salient features about the transition between secondary and tertiary (college) education in the data that we ask our model to replicate. Firstly, the share of children with a academic track secondary school degree who end up getting a college degree is significantly higher than those with a vocational secondary school degree. In Germany, most of this is coming from the fact that an academic track secondary school diploma automatically qualifies for university entrance. We capture this with the two parameters α and $\alpha_{S^c=V}$, which we estimate 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.

Secondly, the likelihood of college education in the data is increasing in the end-of-school skills. 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 selection of higher-skilled school graduates into an academic career, where these (mostly cognitive) skills are more useful. We capture this through the coefficient α_θ, which multiplies end-of-school skills and is expected to be negative. As a target moment for this parameter, we choose the regression coefficient on log math test scores from a regression of a college graduation dummy on end-of-school test scores.

We recognize that there are many pathways into academic higher education in Germany that fall outside the scope of our model.[40] To account for additional heterogeneity in the college decision, we add the normally distributed college taste shock ν, with parental education specific means μ_{ν,S^p} and variance σ_ν^2. We calibrate the two parameters $\mu_{\nu,S^p=V}$ and $\mu_{\nu,S^p=A}$ 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_ν^2 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 are not a part of the "psychic" costs ψ_A but reflect 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.[41] Thus, we set the maximum remaining time during the higher education stage to $\bar{n}(A) = 0.40$.

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, $\{\gamma_{j,S}\}$,

[40] 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.

[41] 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.

$j = 5, ..., 16$ using wage regressions in the SOEP data, following the approach in Lagakos et al. (2018).[42] The resulting experience-wage profiles for 4-year experience bins are shown in Table 2, expressed in growth relative to the previous bin. We set the $\{\gamma_{j,S}\}_{j=5}^{16}$ 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 2: 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

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

[42] 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

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

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).

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, Mayr, and Sachs (2020). The linear capital tax is set to $\tau_s = 0.35$. 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 $\pi_j = \Omega h_j w_S$ during retirement and calibrate the scale parameter Ω internally, such that the average replacement rate corresponds to 40%.

Table 3: 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)
$\bar{n}(S = A)$	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, Mayr, and Sachs (2020)
Government			
τ_n	0.128	Labor Tax Progressivity	Kindermann, Mayr, and Sachs (2020)
λ	0.679	Labor Tax Scale	Kindermann, Mayr, and Sachs (2020)
τ_s	0.35	Capital Tax Rate	
g	0.06	Lump-sum Transfers	

4.3. 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 4.

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

Table 4: Parameters estimated internally

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_V	0.16	Own V-Track Bias	Share of Deviations	0.19	0.19
χ_A	0.17	Own A-Track Bias	Share of Deviations	0.22	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
$\mu_{S^p=V}$	0.18	Mean Taste Shock if $S^p = V$	Share in CL from Non-CL HH	0.20	0.20
$\mu_{\nu,S^p=A}$	-0.18	Mean Taste Shock if $S^p = A$	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	0.418	0.639
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 $j = 4$	$\text{Rank}_{j=3}$-$\text{Rank}_{j=4}$	0.68	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	0.350	0.356

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 college education, while that share is 11% from a vocational track secondary school. Thus, 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 0.42 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.

4.4. 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, Puhani, and Schönberg (2017) for Germany.

Non-targeted Moments

We summarize selected non-targeted moments and their data or external counterparts in Table 5. The first set of moments pertain 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 by parental education decrease slightly during secondary school. In that sense, schooling seems to act as an equalizer as is often argued. While we overestimate the initial difference at the beginning of primary school, our model matches well this trend, both qualitatively and quantitatively.

Similarly, the differences in average child skills across school tracks (in terms of standard deviations) decreases 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 share of college-educated parents who choose an academic track school for their child is 74%. For non-college educated parents, this share is only 24%. The model implies a slight underestimation of the former and overestimation of the later. 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 0.87 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 and underestimates the importance of the parental background and other factors.

The third set of moments relate 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 fifth income rank quintile over first quintile of 2.13, which our model matches well. We also compare our model-implied estimate of the intergenerational elasiticity 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, Puhani, and Schönberg (2017) analyse the long-term labor market effects of early school track choice in Germany using a quasi-experimental setting. Their

Table 5: Non-targeted moments

Moment	Data	Model
Child Skill Moments		
Mean Differences by Parental Background (in Standard Deviations)		
Beginning Secondary School	0.517	0.563
Middle Secondary School	0.434	0.469
End Secondary School	0.379	0.426
Mean Differences by School Track (in Standard Deviations)		
Middle Secondary School	1.020	1.097
End Secondary School	0.775	0.942
School Track Choice		
Share A-track children from CL. HH	0.74	0.68
Share A-track children from Non-CL HH	0.24	0.28
Coefficient A-track on Skill Rank	0.87	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%

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.[43]

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 ($\theta_{j^c=3}$). As detailed in Section 2, our child skill development technology implies that, conditional on

[43]Note that Dustmann, Puhani, and Schönberg (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 will have accumulated less work experience.

parental background, the school track choice is characterized by a skill threshold, such that all children with skills above that threshold go 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 the 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, Puhani, and Schönberg (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 around the track 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 track threshold who go to different school tracks, conditional on all other states.[44] 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.[45] 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 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, Puhani, and Schönberg (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 less differences in lifetime outcomes. In sum, we conclude that the implications our model entails with respect to the effect of tracking on *marginal* children is not at odds with the reduced-form evidence presented in Dustmann, Puhani, and Schönberg (2017).

[44] 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 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^c 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.

[45] 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.

5. 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, Ventura, and Yaron (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 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.

5.1. 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.[46] These states are the school track in secondary school, S^c, initial adult human capital h_5, initial transfers received from the parent s_5, parental education S^p, and innate learning ability ϕ^c.

Table 6: Contributions to Lifetime Inequality

Life Stage	States	Share of Explained Variance Lifetime Earnings	Share of Explained Variance Lifetime Wealth
Independence (age 18)	$(S^c, \phi^c, h_5, s_5, S^p)$	80%	74%
	(S^c, ϕ^c, h_5, S^p)	75%	72%
	(S^c, ϕ^c, s_5, S^p)	57%	44%
School Track Choice (age 10)	$(S^c, \phi^c, \theta_3, h_{11}, s_{11}, S)$	21%	25%
	(S^c, θ_3, ϕ^c)	20%	22%
	(S^c)	15%	19%
Pre-Birth (parent age 30)	(S, ϕ, h_8, s_8)	4.5%	11.3%

[46] 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.

Row 1 of Table 6 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%.[47] 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.

As Row 2 of Table 6 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 s_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 even by almost 30 percentage points. This suggests that variation in human capital, even at age 18, is an important driver of lifetime inequality.[48] 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.[49]

Using the same methodology, we can also evaluate how much of 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 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 4.4, for children with similar skills, the track choice has only small independent effects on

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

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

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

lifetime outcomes. We investigate the determinants and consequences of the school track choice in more detail below.

The last row of Table 6 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 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 an important role of the wealth transfers. For example, if we use the same decomposition of the unconditional variance of transfers into parental states pre-birth, we find that almost 39% of variation 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.

5.2. School Track Choice Counterfactuals

According to the theoretical predictions layed out in Section 2, the initial school track should, to a large degree, be based on selection 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 7 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 7 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, academic (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 her 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,S^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 2, for a set of children with low pre-school 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 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.

Table 7: School Track Choice Determinants

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

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 we isolate each effect, respectively.[50] In particular, we isolate the effects of the first channel by solving the model with $\kappa_{5,j^c=3,4} = 0$ yet leaving $\kappa_{5,j^c=3,4} > 0$ 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 7 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 in 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 suggests 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.

[50]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.

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

The most importance reason for the remaining influence of parental background on a child's school track works through the immediate own-track bias governed by χ_S. As reported in Column (4) of Table 7, the direct influence of parental education on the school track of a child drops by almost 80% if we set $\chi_S = 0$ 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. This suggests that preferences for own-tracks by parents can create inefficiencies in the allocation of children across school tracks. If college educated parents send their children to an academic track school, despite the fact that her skill level would optimally suggest the vocational track, 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 background send their overqualified children there purely based on preferences.

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 8 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, in which 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 2, 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 parental background.[51]

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, 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 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 the counterfactual experiments.

In particular, the first two rows in Panel B. of Table 8 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 $j^c = 4$ 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 7) 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 as in the baseline economy, and prices adjust such that 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

[51] The optimal tracking threshold derived in Section 2 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.

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 income drops in both cases.

Table 8: Effects of School Track Choice Counterfactuals

Outcome	(1) Baseline Economy	(2) $\chi_V = 0$ $\chi_A = 0$	(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%
$Std(\theta_{4\|S^c=V})$	0.21	-1.9%	-5.2%
$Std(\theta_{4\|S^c=A})$	0.33	-0.9%	-2.4%
$Std(\theta_{5\|S^c=V})$	0.28	-0.4%	-1.1%
$Std(\theta_{5\|S^c=A})$	0.28	-0.4%	-1.4%

Overall, the results painted in 8 paint 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 macro-economy. At the same time, both counterfactual experiments result in more upward mobility of children relative to their parents.

5.3. 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 (2013a) 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.

Recall that we assumed that a policymaker sets the pace of instruction, that is the skill level at which children are taught, in each school track optimally with the goal of maximizing learning in that track. As shown before, this gives rise to a skill production technology during school years, in which future skills are a function of the (distance to the) average skill level among all children in a given track. Since we estimated the parameters governing this function separately for each education stage, we do not need to make additional assumptions regarding how a postponed tracking policy reform affects child learning. Instead, (the implicit assumption is that) child learning is affected exclusively through the effects of the late tracking policy on the instruction pace in each track, and hence the average skill level in each track.

We present the relative changes of selected aggregate and social mobility outcomes of the counterfactual experiments relative to the baseline economy in Table 9. The experiments differ in the way we assume that prices and instruction paces are allowed to adjust. In Column (1), all prices (wages per efficiency unit for college and non-college human capital w_V, w_A 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 are unchanged, the college premium that measure earnings divided by labor supply, declines.

As we derived in Section 2, 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 9. 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 9, it is also true that academic track attendance depends even less on the child 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 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 (2), 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 the academic track still increases substantially. Indeed, if we allow for an adjusting instruction pace, still more than 90% of children go 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 1) put less emphasis on the distance to the average peer level compared to $j = 3$. This reduces the learning penalty from sending your child to the academic track.

Both affects are amplified when the teaching pace adjusts endogenously, as less and less 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 in 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 of the parents. Moreover, the track decision seems largely separate from college admission concerns. Third, this

Table 9: Late Tracking Counterfactuals

Outcome	(1) Early Tracking	(2) PE + ET Pace	(3) PE + Pace adjusts	(4) GE + ET Pace	(5) GE + Pace adjusts
			Panel A.		
Y	1.1131	-3.6%	-9.8%	-2.1%	-3.3%
College Share	0.356	48.3%	73.6%	2.2%	4.2%
A-Track Share	0.421	74.6%	131.4%	63.9%	127.6%
College Premium	1.376	-4.0%	-3.8%	-4.8%	-7.5%
Gini Earnings	0.22	0.0%	0.5%	-2.3%	-3.6%
A-Track on Income	0.444	-7.2%	-86.9%	-16.7%	-86.7%
A-Track on Skills	1.418	-44.6%	-90.6%	-40.2%	-85.9%
CL on Skills	0.785	1.9%	-12.2%	-2.5%	-7.3%
IGE	0.278	-8.6%	-14.7%	-8.3%	-11.5%
			Panel B.		
$\bar{\theta}_4$	-0.215	-7.9%	-3.3%	-12.6%	-12.6%
$\bar{\theta}_5$	-0.233	-4.3%	-7.3%	-8.2%	-12.4%
$Std(\theta_{4\|S^c=V})$	0.21	-11.4%	-41.0%	-10.5%	-37.6%
$Std(\theta_{4\|S^c=A})$	0.327	-12.8%	-9.8%	-14.7%	-12.8%
$Std(\theta_{5\|S^c=V})$	0.283	-8.8%	-13.4%	-8.5%	-31.8%
$Std(\theta_{5\|S^c=A})$	0.279	-5.0%	-7.2%	-6.8%	-9.3%

results in a more equal access to academic secondary education, which continuous into higher education and labor market outcomes. It also reduces the persistence of economic status across generations.

6. 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 common policy across many advanced countries, influences not only equality of educational opportunities of 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 schooling years of the children. 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 that is 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 are in line with the most robust findings of a vast empirical literature as well as the most popular arguments in the public debate about tracking. In particular, not every child gains from tracking and the losses are often concentrated among the lower-skilled children. Additionally, tracking can lead to increases 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 age 10 of the child, and calibrate it on German data. Our quantitative results suggest that variation coming 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 own track that cannot be explained by parental inputs into child skills or tastes for higher education. This give rise to efficiency-reducing misallocation of children across tracks. However, 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 macro-economy. 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 a great majority of children being allocated into one track, which makes everyone more equal but harms learning efficiency dramatically.

References

Aakvik, Arild, Kjell G Salvanes, and Kjell Vaage (2010). “Measuring heterogeneity in the returns to education using an education reform”. In: *European Economic Review* 54.4, pp. 483–500.

Abbott, Brant et al. (2019). “Education policy and intergenerational transfers in equilibrium”. In: *Journal of Political Economy* 127.6, pp. 2569–2624.

Agostinelli, Francesco (2018). “Investing in children’s skills: An equilibrium analysis of social interactions and parental investments”. In: *Unpublished Manuscript, University of Pennsylvania*.

Agostinelli, Francesco, Matthias Doepke, et al. (2023). *It takes a village: the economics of parenting with neighborhood and peer effects*. Tech. rep. National Bureau of Economic Research.

Agostinelli, Francesco, Morteza Saharkhiz, and Matthew Wiswall (2019). *Home and School in the Development of Children*. Tech. rep. National Bureau of Economic Research.

Agostinelli, Francesco and Matthew Wiswall (2016). *Estimating the technology of children’s skill formation*. Tech. rep. National Bureau of Economic Research.

Ballou, Dale (2009). “Test scaling and value-added measurement”. In: *Education finance and Policy* 4.4, pp. 351–383.

Becker, Gary S and Nigel Tomes (1986). “Human capital and the rise and fall of families”. In: *Journal of labor economics* 4.3, Part 2, S1–S39.

Bellenberg, Gabriele and Matthias Forell (2012). “Schulformwechsel in Deutschland”. In: *Durchlässigkeit und Selektion in den* 16.

Blossfeld, HP, HG Roßbach, and J von Maurice (2011). “The German National Educational Panel Study (NEPS)”. In: *Zeitschrift für Erziehungswissenschaft: Sonderheft* 14.

Bonesrønning, Hans et al. (2022). “Small-group instruction to improve student performance in mathematics in early grades: Results from a randomized field experiment”. In: *Journal of Public Economics* 216, p. 104765.

Borghans, Lex et al. (2008). “The economics and psychology of personality traits”. In: *Journal of human Resources* 43.4, pp. 972–1059.

Capelle, Damien (2022). *The Great Gatsby goes to College: Tuition, Inequality and Intergenerational Mobility in the U.S.* Tech. rep.

Caucutt, Elizabeth M and Lance Lochner (2020). “Early and late human capital investments, borrowing constraints, and the family”. In: *Journal of Political Economy* 128.3, pp. 1065–1147.

Ciccone, Antonio and Giovanni Peri (2005). “Long-run substitutability between more and less educated workers: evidence from US states, 1950–1990”. In: *Review of Economics and statistics* 87.4, pp. 652–663.

Cunha, Flavio and James Heckman (2007). “The technology of skill formation”. In: *American economic review* 97.2, pp. 31–47.

Cunha, Flavio, James Heckman, and Susanne M Schennach (2010). “Estimating the technology of cognitive and noncognitive skill formation”. In: *Econometrica* 78.3, pp. 883–931.

Daruich, Diego (2022). “The Macroeconomic Consequences of Early Childhood Development Policies”. In: *FRB St. Louis Working Paper* 2018-29.

Dodin, Majed et al. (2021). *Social Mobility in Germany*. Tech. rep. Working Paper.

Duflo, Esther, Pascaline Dupas, and Michael Kremer (2011). “Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya”. In: *American Economic Review* 101.5, pp. 1739–74.

Dustmann, Christian (2004). “Parental background, secondary school track choice, and wages”. In: *Oxford Economic Papers* 56.2, pp. 209–230.

Dustmann, Christian, Patrick A Puhani, and Uta Schönberg (2017). “The long-term effects of early track choice”. In: *The Economic Journal* 127.603, pp. 1348–1380.

Epple, Dennis and Richard Romano (2011). “Peer effects in education: A survey of the theory and evidence”. In: *Handbook of social economics*. Vol. 1. Elsevier, pp. 1053–1163.

Falk, Armin, Fabian Kosse, and Pia Pinger (2021). “Mentoring and schooling decisions: Causal evidence”. In.

Fuchs-Schündeln, Nicola, Dirk Krueger, Andre Kurmann, et al. (2021). *The Fiscal and Welfare Effects of Policy Responses to the Covid-19 School Closures*. Tech. rep. National Bureau of Economic Research.

Fuchs-Schündeln, Nicola, Dirk Krueger, Alexander Ludwig, et al. (2022). “The long-term distributional and welfare effects of Covid-19 school closures”. In: *The Economic Journal* 132.645, pp. 1647–1683.

Fujimoto, Junichi, David Lagakos, and Mitchell Vanvuren (2022). *Aggregate and Distributional Effects of ‘Free’ Secondary Schooling in the Developing World*. Tech. rep. Working Paper.

Guyon, Nina, Eric Maurin, and Sandra McNally (2012). “The effect of tracking students by ability into different schools a natural experiment”. In: *Journal of Human resources* 47.3, pp. 684–721.

Hanushek, Eric A and Ludger Wössmann (2006). “Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries”. In: *The Economic Journal* 116.510, pp. C63–C76.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L Violante (2017). “Optimal tax progressivity: An analytical framework”. In: *The Quarterly Journal of Economics* 132.4, pp. 1693–1754.

Heckman, James and Stefano Mosso (2014). “The economics of human development and social mobility”. In: *Annu. Rev. Econ.* 6.1, pp. 689–733.

Henninges, Miriam, Claudia Traini, and Corinna Kleinert (2020). “Tracking and Sorting in the German Educational System”. In.

Huggett, Mark, Gustavo Ventura, and Amir Yaron (2011). “Sources of lifetime inequality”. In: *American Economic Review* 101.7, pp. 2923–54.

Jang, Youngsoo and Minchul Yum (Nov. 2022). *Aggregate and Intergenerational Implications of School Closures: A Quantitative Assessment*. Discussion Paper Series 234v1. CRC TR 224.

Keane, Michael P and Kenneth I Wolpin (1997). “The career decisions of young men”. In: *Journal of political Economy* 105.3, pp. 473–522.

Kindermann, Fabian, Lukas Mayr, and Dominik Sachs (2020). “Inheritance taxation and wealth effects on the labor supply of heirs”. In: *Journal of Public Economics* 191, p. 104127.

Kyzyma, Iryna and Olaf Groh-Samberg (2018). *» Intergenerational Economic Mobility in Germany: Levels und Trends «*. Tech. rep. DIW Working Paper.

Lagakos, David et al. (2018). “Life cycle wage growth across countries”. In: *Journal of Political Economy* 126.2, pp. 797–849.

Lee, Sang Yoon and Ananth Seshadri (2019). “On the intergenerational transmission of economic status”. In: *Journal of Political Economy* 127.2, pp. 855–921.

Malamud, Ofer and Cristian Pop-Eleches (2011). “School tracking and access to higher education among disadvantaged groups”. In: *Journal of Public Economics* 95.11-12, pp. 1538–1549.

Matthewes, Sönke Hendrik (2021). “Better together? Heterogeneous effects of tracking on student achievement”. In: *The Economic Journal* 131.635, pp. 1269–1307.

Meghir, Costas and Mårten Palme (2005). “Educational reform, ability, and family background”. In: *American Economic Review* 95.1, pp. 414–424.

Neumann, Irene et al. (2013). “Modeling and assessing mathematical competence over the lifespan”. In: *Journal for educational research online* 5.2, pp. 80–109.

OECD (2013a). *PISA 2012 Results: What Makes Schools Successful (Volume IV)*, p. 544.

OECD (2013b). *PISA 2012 Results: What Makes Schools Successful (Volume IV): Resources, Policies and Practices, PISA*. Tech. rep. Expertise im Auftrag der Initiative Neue Soziale Marktwirtschaft.

Pekkala Kerr, Sari, Tuomas Pekkarinen, and Roope Uusitalo (2013). “School tracking and development of cognitive skills”. In: *Journal of Labor Economics* 31.3, pp. 577–602.

Piopiunik, Marc (2014). “The effects of early tracking on student performance: Evidence from a school reform in Bavaria”. In: *Economics of Education Review* 42, pp. 12–33.

Pohl, Steffi and Claus H Carstensen (2012). “NEPS technical report-Scaling the data of the competence tests”. In: *NEPS Working Papers/Survey Papers*.

– (2013). “Scaling of competence tests in the National Educational Panel Study-Many questions, some answers, and further challenges”. In: *Journal for Educational Research Online* 5.2, pp. 189–216.

Ruhose, Jens and Guido Schwerdt (2016). “Does early educational tracking increase migrant-native achievement gaps? Differences-in-differences evidence across countries”. In: *Economics of Education Review* 52, pp. 134–154.

Woessmann, Ludger (2013). “Die entscheidende Säule”. In: *Wirtschaftswoche Global* 2.24.06.2013, pp. 110–111.

– (2016). “The importance of school systems: Evidence from international differences in student achievement”. In: *Journal of Economic Perspectives* 30.3, pp. 3–32.

Yum, Minchul (2022). “Parental time investment and intergenerational mobility”. In: *International Economic Review*.

A. Appendix

A.1. Proof of Propositions

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 $S^c = V$ and all $\theta_1 > \tilde{\theta}_1$ go to the other track, $S^c = A$ (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 (1) of a child in a given school track S^c with instruction pace $P^*_{S^c}$ using Lemma 1 as:

$$\theta_2 = \theta + \alpha\bar{\theta}_{S^c} + \frac{\beta^2}{2\delta} + \frac{\beta\gamma\theta_1}{\delta} + \frac{\gamma^2\theta_1\bar{\theta}_{S_c}}{\delta} - \frac{\gamma^2\bar{\theta}^2_{S_c}}{2\delta} + \eta_2. \tag{A.1}$$

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

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

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 $\bar{\theta}_{S^c}$ 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, $\mathbb{E}(\theta_1|S^c = V) < \mathbb{E}(\theta_1|S^c = A)$. 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 $\theta_1'' \leq \mathbb{E}(\theta_1|S^c = V)$ that goes to track $S^c = A$, and/or a child with $\theta_1'' \geq \mathbb{E}(\theta_1|S^c = A)$ that goes to track $S^c = V$, and/or two children with skills $\theta_1'' < \theta_1'''$, with $\theta_1'', \theta_1''' \in [\mathbb{E}(\theta_1|S^c = V), \mathbb{E}(\theta_1|S^c = A)]$, 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 (A.2), 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^c again indicate to which track a child is allocated, now with $S^c = V$ for all $\theta_1 \leq \tilde{\theta}_1$ and $S^c = A$ for all $\theta_1 > \tilde{\theta}_1$.

A policymaker solves

$$
\begin{aligned}
& \max_{\tilde{\theta}_1} \quad \mathbb{E}(\theta_2) \\
\iff & \max_{\tilde{\theta}_1} \quad \mathbb{E}(\mathbb{E}(\theta_2|S^c)) \\
\text{subject to} & \\
& P_{S^c} \text{ chosen optimally given Lemma 1.}
\end{aligned}
\tag{A.3}
$$

Using (A.1) and the law of iterated expectations, this maximization problem boils down to

$$
\begin{aligned}
& \max_{\tilde{\theta}_1} \quad \frac{\beta^2}{2\delta} + \frac{\gamma^2}{2\delta}\,\mathbb{E}\left(\mathbb{E}(\theta_1|S^c)^2\right) \\
\iff & \max_{\tilde{\theta}_1} \quad \frac{\beta^2}{2\delta} + \frac{\gamma^2}{2\delta}\left(F(\tilde{\theta}_1)\,\mathbb{E}(\theta_1|\theta_1 \leq \tilde{\theta}_1)^2 + (1 - F(\tilde{\theta}_1))\,\mathbb{E}(\theta_1|\theta_1 > \tilde{\theta}_1)^2\right),
\end{aligned}
\tag{A.4}
$$

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)

$$
\begin{aligned}
& \max_{\tilde{\theta}_1} \quad \mathbb{E}(\theta_2) \\
\Longleftrightarrow \quad & \max_{\tilde{\theta}_1} \quad \frac{\gamma^2}{2\delta}\left(\sigma^2_{\theta_1} - \mathbb{E}(Var[\theta_1|S^c])\right).
\end{aligned} \tag{A.5}
$$

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

$$
\begin{aligned}
& \left(\alpha + \hat{\theta}_1\frac{\gamma^2}{\delta}\right)\mathbb{E}(\theta_1|\theta_1 \leq \hat{\theta}_1) - \frac{\gamma^2}{2\delta}\,\mathbb{E}(\theta_1|\theta_1 \leq \hat{\theta}_1)^2 \\
= & \left(\alpha + \hat{\theta}_1\frac{\gamma^2}{\delta}\right)\mathbb{E}(\theta_1|\theta_1 > \hat{\theta}_1) - \frac{\gamma^2}{2\delta}\,\mathbb{E}(\theta_1|\theta_1 > \hat{\theta}_1)^2 \\
\Longleftrightarrow & \left(-\alpha - \hat{\theta}_1\frac{\gamma^2}{\delta}\right)\sigma_{\theta_1}\frac{f(\hat{\theta}_1/\sigma)}{F(\hat{\theta}_1/\sigma)} - \frac{\gamma^2}{2\delta}\sigma^2_{\theta_1}\frac{f(\hat{\theta}_1/\sigma)^2}{F(\hat{\theta}_1/\sigma)^2} \\
= & \left(\alpha + \hat{\theta}_1\frac{\gamma^2}{\delta}\right)\sigma_{\theta_1}\frac{f(\hat{\theta}_1/\sigma)}{1 - F(\hat{\theta}_1/\sigma)} - \frac{\gamma^2}{2\delta}\sigma^2_{\theta_1}\frac{f(\hat{\theta}_1/\sigma)^2}{(1 - F(\hat{\theta}_1/\sigma))^2}
\end{aligned} \tag{A.6}
$$

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 (A.6) numerically. In all cases with reasonable parameter values, (A.6) 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 (A.6). 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

$$\begin{aligned}\mathbb{E}(\theta_{2,V}) &= \frac{\beta^2}{2\delta} + \alpha\bar{\theta}_{1,V} - \frac{\gamma^2}{2\delta}\bar{\theta}^2_{1,V} \\ &= \frac{\beta^2}{2\delta} - \alpha\sigma_{\theta_1}\frac{f(\tilde{\theta}_1/\sigma_{\theta_1})}{F(\tilde{\theta}_1/\sigma_{\theta_1})} - \frac{\gamma^2}{2\delta}\sigma^2_{\theta_1}\frac{f(\tilde{\theta}_1/\sigma_{\theta_1})^2}{F(\tilde{\theta}_1/\sigma_{\theta_1})^2}.\end{aligned} \tag{A.7}$$

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

$$\begin{aligned}\mathbb{E}(\theta_{2,A}) &= \frac{\beta^2}{2\delta} + \alpha\bar{\theta}_{1,A} - \frac{\gamma^2}{2\delta}\bar{\theta}^2_{1,A} \\ &= \frac{\beta^2}{2\delta} + \alpha\sigma_{\theta_1}\frac{f(\tilde{\theta}_1/\sigma_{\theta_1})}{1-F(\tilde{\theta}_1/\sigma_{\theta_1})} - \frac{\gamma^2}{2\delta}\sigma^2_{\theta_1}\frac{f(\tilde{\theta}_1/\sigma_{\theta_1})^2}{(1-F(\tilde{\theta}_1/\sigma_{\theta_1}))^2}.\end{aligned} \tag{A.8}$$

The variance of θ_2 in a comprehensive system is

$$\begin{aligned}Var(\theta_{2,C}) &= \mathbb{E}(\big(\theta_2 - \mathbb{E}(\theta_2))^2\big) \\ &= (1+\frac{\beta\gamma}{\delta})^2\sigma^2_{\theta_1} + \sigma^2_{\eta_2} \\ &\sigma^2_{\theta_2,C} + \sigma^2_{\eta_2},\end{aligned} \tag{A.9}$$

where we define $\sigma^2_{\theta_2,C}$ 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

$$\begin{aligned}Var(\theta_{2,V}) &= \mathbb{E}\left(\theta_2 - \mathbb{E}(\theta_{2,V})\right)^2 \\ &= \mathbb{E}\left(\theta_1 + \alpha\bar{\theta}_{1,V} + \frac{\beta^2}{2\delta} + \frac{\beta\gamma}{\delta}\theta_1 + \frac{\gamma^2\theta_1\bar{\theta}_{1,V}}{\delta} - \frac{\gamma^2\bar{\theta}^2_{1,V}}{2\delta} + \eta_2 - (\alpha\bar{\theta}_{1,V} + \frac{\beta^2}{2\delta} - \frac{\gamma^2\bar{\theta}^2_{1,V}}{2\delta})\right)^2 \\ &= \sigma^2_{\theta_1}\left[(1+\frac{\beta\gamma}{\delta})^2 + 2(1+\frac{\beta\gamma}{\delta})\frac{\gamma^2}{\delta}\bar{\theta}_{1,V} + \frac{\gamma^4}{\delta^2}\bar{\theta}^2_{1,V}\right] + \sigma^2_{\eta_2} \\ &= \sigma^2_{\theta_2,C} + \sigma^2_{\theta_1}\left[2(1+\frac{\beta\gamma}{\delta})\frac{\gamma^2}{\delta}\bar{\theta}_{1,V} + \frac{\gamma^4}{\delta^2}\bar{\theta}^2_{1,V}\right] + \sigma^2_{\eta_2} = \sigma^2_{\theta_{2,V}} + \sigma^2_{\eta_2}\end{aligned} \tag{A.10}$$

and similarly for track A

$$Var(\theta_{2,A}) = \sigma^2_{\theta_2,C} + \sigma^2_{\theta_1}\left[2(1+\frac{\beta\gamma}{\delta})\frac{\gamma^2}{\delta}\bar{\theta}_{1,A} + \frac{\gamma^4}{\delta^2}\bar{\theta}^2_{1,A}\right] + \sigma^2_{\eta_2} = \sigma^2_{\theta_{2,A}} + \sigma^2_{\eta_2}, \tag{A.11}$$

where, again we define $\sigma^2_{\theta_2,S^c}$ to be the unconditional variance in track S^c net of the skill shock variance.

Given this, the conditional expected value of θ_2 in track S^c, that is the variance among those actually in that track is a truncation of the unconditional normally distributed θ_{2,S^c}, 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 θ_{2,S^c} as expressed above. Thus, we can find the conditional expected value of skills in track S^c using the formula for a truncated normal distribution and the computed unconditional standard deviation of θ_2 in each track,, $\sigma_{\theta_{2,S^c}}$, *as if* there were no skill shock realizations which yields

$$\begin{aligned}\mathbb{E}(\theta_{2,V}|S^c = V) &= \mathbb{E}(\theta_{2,V}) - \sigma_{2,V}\frac{f(\tilde{\theta}_1/\sigma_{\theta_2,V})}{F(\tilde{\theta}_1/\sigma_{\theta_2,V})} \\ &= \frac{\beta^2}{2\delta} + \alpha\bar{\theta}_{1,V} - \frac{\gamma^2}{2\delta}\bar{\theta}^2_{1,V} - \sigma_{\theta_2,V}\frac{f(\tilde{\theta}_1/\sigma_{\theta_2,V})}{F(\tilde{\theta}_1/\sigma_{\theta_2,V})},\end{aligned} \tag{A.12}$$

for the V track and

$$\begin{aligned}\mathbb{E}(\theta_{2,A}|S^c = A) &= \mathbb{E}(\theta_{2,A}) + \sigma_{2,A}\frac{f(\tilde{\theta}_1/\sigma_{2,A})}{1 - F(\tilde{\theta}_1/\sigma_{\theta_2,A})} \\ &= \frac{\beta^2}{2\delta} + \alpha\bar{\theta}_{1,A} - \frac{\gamma^2}{2\delta}\bar{\theta}^2_{1,A} + \sigma_{\theta_2,A}\frac{f(\tilde{\theta}_1/\sigma_{\theta_2,A})}{1 - F(\tilde{\theta}_1/\sigma_{\theta_2,A})},\end{aligned} \tag{A.13}$$

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

$$Var(\theta_{2,V}|S^c = V) = \sigma^2_{\theta_{2,V}}\left(1 - \frac{\tilde{\theta}_1}{\sigma_{\theta_{2,V}}}\frac{f(\tilde{\theta}_1/\sigma_{\theta_2,V})}{F(\tilde{\theta}_1/\sigma_{\theta_2,V})} - \frac{f(\tilde{\theta}_1/\sigma_{\theta_2,V})^2}{F(\tilde{\theta}_1/\sigma_{\theta_2,V})^2}\right) + \sigma^2_{\eta_2},$$

$$Var(\theta_{2,A}|S^c = A) = \sigma^2_{\theta_{2,A}}\left(1 + \frac{\tilde{\theta}_1}{\sigma_{\theta_{2,A}}}\frac{f(\tilde{\theta}_1/\sigma_{\theta_2,A})}{1 - F(\tilde{\theta}_1/\sigma_{\theta_2,A})} - \frac{f(\tilde{\theta}_1/\sigma_{\theta_2,A})^2}{(1 - F(\tilde{\theta}_1/\sigma_{\theta_2,A}))^2}\right) + \sigma^2_{\eta_2}.$$

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

$$Var(\theta_{2,T}) = \mathbb{E}(Var(\theta_{2,S^c}|S^c)) + Var(\mathbb{E}(\theta_{2,S^c}|S^c)). \tag{A.14}$$

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

$$\mathbb{E}(Var(\theta_{2,S^c}|S^c)) = \sigma^2_{\eta_2} + (1 - \frac{f(0)^2}{0.25}) \left(\sigma^2_{\theta_{2,C}} + \sigma^4_{\theta_1}\frac{\gamma^4}{\delta^2}\frac{f(0)^2}{0.25}\right). \tag{A.15}$$

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

$$\begin{aligned} Var(\mathbb{E}(\theta_{2,S^c}|S^c)) &= \left(\alpha\bar{\theta}_{1,V} - f(0)(\sigma_{2,v} + \sigma_{2,A})\right)^2 \\ &= \left(\alpha\bar{\theta}_{1,A} + f(0)(\sigma_{2,v} + \sigma_{2,A})\right)^2. \end{aligned} \tag{A.16}$$

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 $\tilde{\theta}_1 = 0$

$$\begin{aligned} Var(\theta_{2,T}) &= \sigma^2_{\eta_2} + (1-\chi)\sigma^2_{\theta_{2,C}} + \chi\frac{\gamma^4}{\delta^2}(2-4\chi)\sigma^4_{\theta_1} \\ &+ 2\chi\alpha\sigma_{\theta_1}(\alpha\sigma_{\theta_1} + \sigma_{\theta_{2,V}} + \sigma_{\theta_{2,A}}) + \chi\sigma_{\theta_{2,V}}\sigma_{\theta_{2,A}} \\ &= \sigma^2_{\eta_2} + \sigma^2_{\theta_{2,C}} + 2\chi\sigma^2_{\theta_1}\left(\alpha^2 + 2\alpha(1 + \frac{\beta\gamma}{\delta}) + \frac{\gamma^4}{\delta^2}\sigma^2_{\theta_1}\right) \\ &- \left(4\chi\frac{\gamma^2}{\delta}\sigma^2_{\theta_1}\right)^2, \end{aligned} \tag{A.17}$$

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

$$\begin{aligned} & Var(\theta_{2,T}) - Var(\theta_{2,C}) \geq 0 \\ \iff \quad & \alpha^2 + 2\alpha\left(1 + \frac{\beta\gamma}{\delta}\right) - (8-\pi)\frac{\gamma^4}{\pi\delta^2}\sigma^2_{\theta_1} \geq 0. \end{aligned} \tag{A.18}$$

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 (A.1)), 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, (A.2) 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

$$\begin{aligned} & \theta_{1,L} + \alpha\bar{\theta}_{1,C} + \frac{\beta^2}{2\delta} + \frac{\beta\gamma}{\delta}\theta_{1,L} + \frac{\gamma^2}{\delta}\bar{\theta}_{1,C}\theta_{1,L} - \frac{\gamma^2}{2\delta}\bar{\theta}_{1,C}^2 + \eta_2 \\ = \; & \theta_{1,L} + \alpha\bar{\theta}_{1,V} + \frac{\beta^2}{2\delta} + \frac{\beta\gamma}{\delta}\theta_{1,V} + \frac{\gamma^2}{\delta}\bar{\theta}_{1,V}\theta_{1,L} - \frac{\gamma^2}{2\delta}\bar{\theta}_{1,V}^2 + \eta_2 \qquad \text{(A.19)} \\ \iff \quad \theta_{1,L} = \; & \frac{1}{2}\bar{\theta}_{1,V} - \frac{\alpha\delta}{\gamma^2}. \end{aligned}$$

Similarly, the upper intersection is given at

$$\theta_{1,U} = \frac{1}{2}\bar{\theta}_{1,A} - \frac{\alpha\delta}{\gamma^2}. \qquad \text{(A.20)}$$

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 $\frac{\gamma^2}{2\delta}\bar{\theta}_{1,V}^2 = \frac{\gamma^2}{2\delta}\bar{\theta}_{1,A}^2$.

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

$$\begin{aligned}
\mathbb{E}(\theta_{3,LT}) &= \mathbb{E}(\mathbb{E}(\theta_{3,LT}|S^2_{LT})) \\
&= \mathbb{E}(\theta_{2,LT}) + \frac{\beta^2}{2\delta} + (\alpha + \frac{\beta\gamma}{\delta})\,\mathbb{E}(\mathbb{E}(\theta_{2,LT}|S^c_{LT})) + \frac{\gamma^2}{2\delta}\,\mathbb{E}(\mathbb{E}(\theta_{2,LT}|S^c_{LT})^2) \\
&= (2 + \alpha + \frac{\beta\gamma}{\delta})\frac{\beta^2}{2\delta} + \frac{\gamma^2}{2\delta}[\sigma^{\theta_2,LT} - \mathbb{E}(Var(\theta_{2,LT}|S^c_{LT}))],
\end{aligned} \tag{A.21}$$

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 (A.9)). The variable S^c_{LT} indicates the track selection in period 2, which follows the cut-off rule $S^c_{LT} = V$ if $\theta_{2,LT} \leq \tilde{\theta}_{2,LT}$ and $S^c_{LT} = A$ otherwise. The cut-off that maximizes (A.21) is $\tilde{\theta}^*_{2,LT} = \mathbb{E}(\theta_{2,LT}) = \frac{\beta^2}{2\delta}$. This follows as (A.21) 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

$$\begin{aligned}
\mathbb{E}(\theta_{3,ET}) &= \mathbb{E}(\mathbb{E}(\theta_{3,ET}|S^2_{ET})) \\
&= \frac{\beta^2}{2\delta} + (1 + \alpha + \frac{\beta\gamma}{\delta})\,\mathbb{E}(\mathbb{E}(\theta_{2,ET}|S^c_{ET})) + \frac{\gamma^2}{2\delta}\,\mathbb{E}(\mathbb{E}(\theta_{2,ET}|S^c_{ET})^2) \\
&= \frac{\beta^2}{2\delta} + (1 + \alpha + \frac{\beta\gamma}{\delta})\left(\frac{\beta^2}{2\delta} + \frac{\gamma^2}{2\delta}[\sigma^2_{\theta_1} - \mathbb{E}(Var(\theta_{1,ET}|S^c_{ET}))]\right) + \frac{\gamma^2}{2\delta}\,\mathbb{E}(\mathbb{E}(\theta_{2,ET}|S^c_{ET})^2 \\
&= \frac{\beta^2}{2\delta} + (1 + \alpha + \frac{\beta\gamma}{\delta})\left(\frac{\beta^2}{2\delta} + \frac{\gamma^2}{2\delta}[\sigma^2_{\theta_1} - \mathbb{E}(Var(\theta_{1,ET}|S^c_{ET}))]\right) \\
&+ \frac{\gamma^2}{2\delta}[\sigma^2_{\theta_2,ET} - \mathbb{E}(Var(\theta_{2,ET}|S^c_{ET}))].
\end{aligned} \tag{A.22}$$

Comparing (A.21) and (A.22), the condition that governs if average end-of-school skills in a late tracking system are larger than in an early tracking system reads

$$\begin{aligned}
&\mathbb{E}(\theta_{3,LT}) - \mathbb{E}(\theta_{3,ET}) \\
&= \frac{\gamma^2}{2\delta}\left(\mathbb{E}(\mathbb{E}(\theta_{2,LT}|S^c_{LT})^2) - \mathbb{E}(\mathbb{E}(\theta_{2,ET}|S^c_{ET})^2)\right) \\
&- (1+\alpha+\frac{\beta\gamma}{\delta})\frac{\gamma^2}{\delta}\,\mathbb{E}(\mathbb{E}(\theta_1|S^c_{ET})^2) > 0.
\end{aligned} \tag{A.23}$$

The last term of (A.23) 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

$$\mathbb{E}(\theta_{2,LT}|S^c_{LT}=V) = \frac{\beta^2}{2\delta} - \sigma_{\theta_2,LT}\frac{f(\tilde{\theta}_{2,LT}/\sigma_{\theta_2,LT})}{F(\tilde{\theta}_{2,LT}/\sigma_{\theta_2,LT})} \tag{A.24}$$

and

$$\mathbb{E}(\theta_{2,LT}|S^c_{LT}=A) = \frac{\beta^2}{2\delta} + \sigma_{\theta_2,LT}\frac{f(\tilde{\theta}_{2,LT}/\sigma_{\theta_2,LT})}{1-F(\tilde{\theta}_{2,LT}/\sigma_{\theta_2,LT})}, \tag{A.25}$$

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 (A.9). 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 (A.12) and (A.13) depend on $\sigma^2_{\theta_{2,S^c}}$, which as defined in (A.10) and (A.11) do not depend on the skill shock variance.

Condition (A.23) 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:

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

Condition (A.23) then becomes

$$
\begin{aligned}
&\mathbb{E}(\theta_{3,LT}) - \mathbb{E}(\theta_{3,ET}) \\
&= \frac{\gamma^2}{2\delta}\Bigg(2\chi\sigma^2_{\eta_2} - 2\chi\sigma^2_{\theta_1}\Bigg(\alpha^2 + \frac{\gamma^4}{\delta^2}f(0)^2\sigma^2_{\theta_1} - \frac{\beta^2\gamma^2}{2\delta^2} \\
&+ \frac{\beta^2\gamma^2}{\delta^2} + 2\alpha(1+\frac{\beta\gamma}{\delta}) - 4\frac{\gamma^4}{\delta^2}f(0)^2\sigma^2_{\theta_1} + 2\chi\frac{\gamma^4}{\delta^2}\sigma^2_{\theta_1} + 1 + \alpha + \frac{\beta\gamma}{\delta}\Bigg)\Bigg) \\
&= \frac{\gamma^2}{\pi\delta}\Bigg(\sigma^2_{\eta_2} - \sigma^2_{\theta_1}\Bigg(1 + \alpha + \alpha^2 + \frac{\beta\gamma}{\delta} + \frac{\beta^2\gamma^2}{2\delta^2} + 2\alpha(1+\frac{\beta\gamma}{\delta}) + \frac{\gamma^4}{2\delta^2\pi}\sigma^2_{\theta_1}\Bigg)\Bigg) > 0.
\end{aligned}
\tag{A.26}
$$

From this, Proposition 3 follows.

A.2. Empirical Evidence on School Track Selection

In this section, we present reduced-form evidence on the effect of parental background on the school track choice for their children. For that we use data from the NEPS Starting Cohort 3. For this cohort, we have information on the school track a child attended in grade 5 as well as the school track that was recommended to the child by her primary school teacher. Moreover, we have information on the highest education level of the parents. We define the dummy variable High SES as being one whenever at least one of the parents obtained college education. Finally, we have information about the test scores each child achieved at the very beginning of secondary school (in grade 5) and towards the end of secondary school (grade 9). See also Section A.5 for more details on the test scores.

The left panel of Table A.1 shows the results of a regression an academic track dummy on the High SES dummy, controlling for age and gender of the child. In the first column, we do not control for math test scores at the beginning of secondary school. The estimated coefficient of 0.35 suggests that, unconditionally, children from parents with a college background are 35 percentage points more likely to go to an academic track school. In the second column, we control for the math test scores of the child. Even then, children from a higher education family background are significantly more likely to go to an academic track school. The conditional SES gap in academic track attendance is 24 percentage points. Both the unconditional and the conditional SES gap in the NEPS data are consistent with other estimates from Germany, such as Falk, Kosse, and Pinger (2021).

The right panel compares the reaction of high and low SES parents to the track recommendations of their child. For example, the top two rows suggest that, after having received an academic track recommendation, almost 95% of high SES parents follow that recommendation and send their child to an academic track. However, among low SES parents, this share is only 81%. Thus, a significant portion of low SES parents deviate downwards from the recommended track. Vice versa, after receiving a vocational track recommendation, more than 20% of high SES parents deviate from that recommendation and send their child the academic track regardless. This share is less than 10% for the low SES parents. Reasons for this apparent own-track bias are discussed in Section 2.

Table A.1: Parental Influence in School Track Selection

	Academic Track	
High SES	0.35	0.24
	(0.02)	(0.02)
Controls:		
Age & Gender	yes	yes
Tests	no	yes
R^2	0.2	0.36
N	2,480	2,475

Deviations from Teacher Recom.	High SES	Low SES
Academic Recom.		
Follow	94%	81%
Deviate	6%	19%
Vocational Recom.		
Follow	78%	91%
Deviate	22%	9%

In Table A.2, we present the results of a linear regression of test scores (math and reading) in grade 9 on test scores at the beginning of secondary school (grade 5), an academic track dummy, a parental high SES dummy and dummies that equal one if the track choice deviated up (that is from vocational recommendation to academic actual track) or down (from academic recommendation to vocational actual track). The regressions also control for age and gender of the child.

The estimated coefficient on the upward deviation are negative and statistically significant. The size of the coefficient suggest that, when math score is the dependent variable, having deviated up undoes around half of the learning game that comes from being in an academic track. In terms of reading outcomes, the deviation up dummy even completely undoes the gains associated from being in an academic track.

Similarly, the coefficient on downward deviation is positive. It is statistically significant only when reading scores are the dependent variable. Its size suggests that less than half of the average loss from not being in an academic track school is undone because the child deviated downward.

In sum, these results suggest firstly that track recommendations are, on average, reasonable. That is because deviations from the recommendations are harmful in terms of learning outcomes of the child in case of downward deviations and inconsequential for children in case of upward deviations.

Table A.2: Effect of Deviations on Learning

	Dependent Variable	
Coefficient	Math Score (Grade 9)	Reading Score (Grade 9)
Score (Grade 5)	0.574	0.427
	(0.023)	(0.020)
Ac. Track	0.101	0.106
	(0.011)	(0.010)
Deviation Up	-0.054	-0.100
	(0.019)	(0.020)
Deviation Down	0.027	0.040
	(0.018)	(0.020)
High SES	0.028	0.028
	(0.009)	(0.009)
R^2	0.456	0.387
N	1,904	1,816

A.3. Equilibrium Definition

We introduce some notation to define the equilibrium more easily. Let $x_j \in X_j$ 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 $\{d^S(x_5)\}$, for school track $\{d^{S^c}(x_{11})\}$, consumption, labor supply, and assets holdings $\{c_j(x_j), n_j(x_j), s_j(x_j)\}$, and parental transfers $\{s_5\,(x_j)\}$; value functions $\{V_j\,(x_j)\}$; (iii) aggregate capital and labor inputs $\{K, H_V, H_A\}$; (iv) prices $\left\{r, w^V, w^A\right\}$; and (v) average skill levels among children in school track S^c $\{\bar{\theta}_{j,S^c}\}$ 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:

$$H_V = \sum_{j=5}^{J_r} \int_{X_j} n_j(x_j)\ h_j\,(x_j)\, d\Theta(X \mid S = V) + \sum_{j=5}^{5} \int_{X_j} n_j(x_j)\ h_j\,(x_j)\, d\Theta(X \mid S = A)$$

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

For college level:

$$H_A = \sum_{j=6}^{J_r} \int_{X_j} n_j(x_j)\, h_j\,(x_j)\, d\Theta(X \mid S = A).$$

5. Asset market clears

$$K = \sum_{j=J_e}^{J_d} \int_{X_j} s_j(x_j) d\Theta(X),$$

which implies that the goods market clears;

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

A.4. 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, Traini, and Kleinert (2020). Figure A.1 illustrates a simplified structure of the system, starting in Grade 4 and ending with tertiary education.

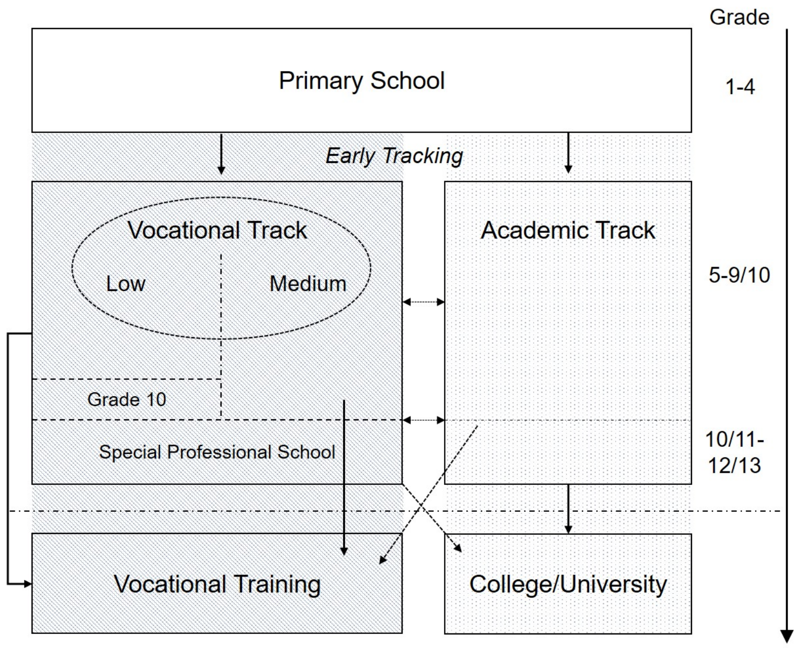
Generally, schooling is mandatory in Germany for every child starting at age six and lasting for nine or ten years. At age six, all children visit a comprehensive primary school that last the first four grades.[52] After that, children are allocated into traditionally three different secondary school 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 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

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

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 A.1: Simplified Structure of the German Education System



The public expenditure per student does not differ significantly across school tracks. Table A.3 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.

A remaining driver behind achievement differences across school tracks could be the teaching quality. In particular higher-quality teachers could select into 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.

Table A.3: 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.5. Measuring Child Skills in the NEPS

In this section, we provide an overview about 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 question, as well as short-constructed responses.[53]

In order to use these questions for the analysis of latent competencies, they need to be scaled. The NEPS (similar to the PISA) using a scarling 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

[53] A simple multiple choice question consists of one correct out of four answer categories, 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 consist of simple multiple choice questions.

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 right answers given to a set of questions by a number of respondents contains 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:

$$p(X_{vi} = 1) = 1 - p(X_{vi} = 0) = \frac{\exp(\theta_v - \sigma_i)}{1 + \exp(\theta_v - \sigma_i)}, \tag{A.27}$$

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.[54] This model is estimated via (weighted) conditional maximum likelihood under a normality assumption on latent ability.

A.6. Details on Child Skill Technology Estimation

In this Section, we provide details on the instrumental variables and data used to estimate the child skill formation technology. We estimate (23) using the (IRT) measure of mathematics tests. This is because we have this measure available at every stage j^c. 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 $j^c = 2$ 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, $j^c = 3$, 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 latter case.

The third stage corresponds to the fourth period in a child's life, $j^c = 4$, when they are between 14/15 and 18 years old and typically finish secondary school. We use the NEPS

[54]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.

Starting Cohort 4 to estimate the technology parameters in this case, again relying on transformed math test scores in grade 9 and grade 12.[55] 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 8 children. This is because it is not uncommon that some primary schools, especially in rural areas feature quite small class sizes in Germany, whereas class sizes are typically in the range of 20-30 in secondary school.

A.7. Supplementary Tables

[55]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 the are 18 years old, continue to learn in an environment that is the same as if 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.

Table A.4: 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)
$\hat{\kappa}_{2,j}$	$\theta^2_{i,j}$	-	0.162	0.189	0.137
		-	(0.115)	(0.110)	(0.070)
$\hat{\kappa}_{3,j}$	$\bar{\theta}_{-i,C_{S(i),j}}$	-0.020	0.145	0.196	0.054
		(0.047)	(0.099)	(0.081)	(0.068)
$\hat{\kappa}_{4,j}$	$(\theta_{i,j} - \bar{\theta}_{S(i),j})^2$	-0.088	-0.229	-0.304	-0.193
		(0.047)	(0.126)	(0.124)	(0.081)
$\hat{\kappa}_{5,j}$	$S = A$	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 A.5: 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
$\hat{\kappa}_{1,j}$	$\ln\theta_{i,j}$	0.577	0.558	0.598	0.521
		(0.016)	(0.025)	(0.024)	(0.021)
$\hat{\kappa}_{2,j}$	$\ln\theta_{i,j}^2$	-	0.174	0.197	0.142
		-	(0.115)	(0.110)	(0.070)
$\hat{\kappa}_{3,j}$	$\ln\bar{\theta}_{-i,C_{S(i)},j}$	-	0.569	0.419	0.166
		-	(0.194)	(0.095)	(0.136)
$\hat{\kappa}_{4,j}$	$(\ln\theta_{i,j} - \ln\bar{\theta}_{S(i),j})^2$	-0.088	-0.240	-0.316	-0.199
		(0.047)	(0.127)	(0.122)	(0.081)
$\hat{\kappa}_{5,j}$	$S=A$	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 A.6: 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
$\hat{\kappa}_{1,j}$	$\ln \theta_{i,j}$	0.956	1.132	0.899	0.778
		(0.027)	(0.136)	(0.062)	(0.055)
$\hat{\kappa}_{2,j}$	$\ln \theta_{i,j}^2$	-	0.333	0.591	0.308
		-	(0.317)	(0.277)	(0.157)
$\hat{\kappa}_{3,j}$	$\ln \bar{\theta}_{S(i),j}$	-	-26.049	5.696	-0.422
			(27.718)	(6.825)	(4.080)
$\hat{\kappa}_{4,j}$	$(\ln \theta_{i,j} - \ln \bar{\theta}_{S(i),j})^2$	-0.211	-0.621	-0.921	-0.591
		(0.136)	(0.549)	(0.411)	(0.211)
$\hat{\kappa}_{5,j}$	$S = A$	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