I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13387

Mentoring and Schooling Decisions: Causal Evidence

Armin Falk
Fabian Kosse
Pia Pinger

JUNE 2020

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13387

Mentoring and Schooling Decisions: Causal Evidence

Armin Falk
briq, University of Bonn and IZA

Fabian Kosse
LMU Munich, briq and IZA

Pia Pinger
University of Cologne, briq and IZA

JUNE 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Mentoring and Schooling Decisions: Causal Evidence*

Inequality of opportunity strikes when two children with the same academic performance are sent to different quality schools because their parents differ in socio-economic status. Based on a novel dataset for Germany, we demonstrate that children are significantly less likely to enter the academic track if they come from low socio-economic status (SES) families, even after conditioning on prior measures of school performance. We then provide causal evidence that a low-intensity mentoring program can improve long-run education outcomes of low SES children and reduce inequality of opportunity. Low SES children, who were randomly assigned to a mentor for one year are 20 percent more likely to enter a high track program. The mentoring relationship affects both parents and children and has positive long-term implications for children's educational trajectories.

JEL Classification: C90, I24, J24, J62

Keywords: mentoring, childhood intervention programs, education, human capital investments, inequality of opportunity, socio-economic status

Corresponding author:
Armin Falk
Institute on Behavior and Inequality (briq)
Schaumburg-Lippe-Straße 5-9
D-53113 Bonn
Germany
E-mail: armin.falk@uni-bonn.de

* Financial support through the Eleven gGmbH, the German Research Foundation (DFG) through Leibniz Programme CRC TR 190, CRC TR 224, and Germany's Excellence Strategy (EXC 2126/1-390838866) is gratefully acknowledged. A previous version of this research was titled "(In)Equality of Opportunity, Mentoring, and Critical Educational Decisions".

1 Introduction

Rising inequalities are a major political concern in many Western societies. An important cause, and manifestation, of societal inequality is unequal access to high quality education and its subsequent consequences. In this paper, we therefore ask whether childhood intervention programs such as mentoring can improve educational outcomes for disadvantaged children, and thereby abate educational disparities. We provide an affirmative, causal answer to this question, adding to a growing literature showing that early childhood interventions can effectively level the playing field and reduce inequalities (e.g., Currie, 2001; Garces et al., 2002; Heckman et al., 2013; Kautz et al., 2014; Fryer et al., 2015; List et al., 2018; Cappelen et al., 2020).

A particularly explicit form of differential access to education is early tracking, a feature that is present in many educational systems. For example, among the 36 OECD countries, 34 have a tracking system whereby children at various ages are selected into more or less academic school types (see table A1).[1] Different tracks are typically associated with varying levels of educational quality and predictive of life outcomes.[2] Being enrolled in a "lower" track often implies lower teacher quality, fewer total hours of instruction, less academic subjects being taught, fewer high-potential peer groups and generally a lower probability to qualify for high-skilled jobs, compared to being enrolled in "higher" tracks. Given the economic and social importance of tracking, the selection process into different tracks imparts information about a society's level of equality of opportunity: Selection can be primarily based on merits such as performance, achievement and motivation. In contrast, in less equal societies, selection is significantly determined by socio-economic status and parental background, i.e., features beyond an individual's control. In the latter case, the educational system may actually enlarge pre-existing inequalities, rather than reducing them.

In the following, we study inequality implications of the tracking process in the context of the German school system, where tracking takes place after elementary school, i.e., at around age 10.[3] We first explore to what extent selection into aca-

[1] We use the term tracking for school systems in which students are bifurcated into different schools, with either a vocational or academic emphasis. Our main focus thus differs from within school classroom allocation based on achievement as seen in the United States or Canada (see Betts, 2011, for a discussion).

[2] E.g., Hanushek and Wössmann (2006) show that tracking increases educational inequality. Figures A1 and A2 in the Appendix show large and lasting differences regarding health and happiness between graduates from different school tracks in Germany.

[3] The German context is particularly interesting to study tracking, as it combines largely public and tuition-free schools and universities with restrictive early tracking. There are few financial constraints to prevent students from obtaining university education but entering university requires

demic (high) or vocational (low) track programs reflects merit versus socio-economic status (SES) and parental background. Building on these findings, we then ask the question whether and how a randomized childhood intervention can reduce observed inequalities and help improve equality of opportunity.

Our data come from the briq family panel (bfp) (for an overview see Falk and Kosse, 2020), which combines comprehensive yearly interviews of children and their families with a randomized intervention in form of a one-to-one mentoring program implemented at the start of the panel in 2011. The panel consists of more than 700 high and low SES families with children born between September 2002 and August 2004. Participation in the mentoring program was randomly assigned among children from low SES families. The mentoring program is a well-established nonprofit program called Balu und Du (German for "Baloo and You"). It provides children with a volunteer mentor, typically a university student, for the duration of one year. Conceptually, the program endeavors to extend a child's horizons and foster the acquisition of new skills and experiences through social interactions between mentor and child. Before and after the intervention, as well as in several follow-up data collections, children and their parents were interviewed by trained interviewers. The dataset includes not only detailed schooling-related information such as grades and attended school track, but also information regarding the tracking-related decision-making process, allowing us to zoom in on the mechanisms of this process.

Our main findings can be summarized as follows. We first document that a child's socio-economic background is an important determinant of track selection, i.e., inequalities in opportunity arise from early tracking. In particular, children from high SES families are significantly more likely to attend the high track than children from low SES backgrounds. The respective conditional SES gap amounts to 31.1 percentage points, and is significant at any conventional level. Relative to a baseline probability of 48.8 percent of attending the high track among low SES children, this is a sizable effect. One could argue that this unconditional, raw effect overstates the extent to which SES differences reflect inequality of opportunity, simply because some differences may arise in response to different levels of effort or motivation rather than parental background. For instance, if low SES students work less hard, achievement gaps do not necessarily reflect inequality of opportunity, but differences in work effort. We therefore also study the "conditional" gap, which describes SES-differences conditional on GPA in elementary school as well as sex and age. This way, the conditional gap expresses any differences that arise

graduation from the academic high school track. Tracking after elementary school is thus a critical junction in children's educational careers.

from SES differences after netting out performance differences. Note that while the unconditional gap possibly overstates inequality of opportunity, the conditional gap understates it if school performance itself is determined by SES differences, e.g. through children's differential access to high-quality learning environments. As expected, the conditional SES gap is smaller than the unconditional effect but with 21.7 percentage points still sizable, and significant at any conventional level. In other words, even conditional on GPA, sex and age, children from low SES backgrounds are significantly less likely to enter the high track. In fact, accounting for performance differences explains less than one third of the unconditional gap in high track attendance. Moreover, the track selection pattern turns out to be very stable over time. Five to six years after tracking took place, both gaps, the conditional as well as the unconditional, remain present and even slightly increased. The apparent irreversibility of track choice further underscores the importance of early tracking.

Building on our first result, our main interest concerns the question whether childhood intervention programs can help reduce the SES gap in early tracking. For that matter, we compare tracking outcomes of untreated and treated low SES children. The latter were randomly assigned to the mentoring program in grades two or three, that is, one to two years prior to the point of tracking. We hypothesized that participation in the mentoring program can affect tracking decisions for various reasons. First, the mentoring program stimulates so-called informal learning, i.e., it is meant to support children in their psycho-social development by widening horizons and fostering self-confidence. As a consequence, treated children might become more motivated and confident in the school context. If so, teachers who spend four years with their students and know them well should attest that treated children have a higher potential for a successful high track career in comparison to untreated children. Second, mentors are important role models (Rhodes et al., 2000; Kosse et al., 2020). In the given context, this is of particular importance because all mentors have completed the high track in school and were enrolled at some college or university. Thereby, the mentor as a role model introduces the notion of high track attendance, and higher education in general, to a family context that is often rather unfamiliar with these concepts. It is therefore likely that children are motivated and encouraged to "imitate" the mentor. In addition, parents are made familiar with high track education and mentors may directly or indirectly motivate parents to consider a high track education for their children.[4] In sum, we expect a positive effect of mentoring on the likelihood of entering high track education. In terms of

[4]See, e.g., Goodman et al. (2019) for a related mechanism, i.e., regarding sibling spillovers in college enrollment.

measures, we further expect that treated children are more likely to receive high track recommendations (child effect) and that parents are more likely to overrule a low track recommendation (parental effect).

Our results confirm these hypotheses. Most importantly, we find a positive causal effect of mentoring on high track attendance among low SES students. The effect amounts to 11 percentage points ($p < 0.05$), i.e., the mentoring intervention closes about one third of the unconditional SES gap. In terms of the conditional difference (accounting for GPA, sex and age), the gap is even reduced by roughly one half. Importantly, the positive effect of mentoring is not short-lived but remains virtually unchanged five to six years after the tracking decision had taken place. The enduring nature of the treatment effect is informative also for the question whether the treatment effect pushed the "wrong" children into the high track. Are children who would have been better off in the low track now urged to enter the high track, with negative consequences? This seems not to be the case. First, we do not observe treatment specific reversals of the tracking decision, nor increases in grade retention. Moreover, we find no differences between treatment and control when asking about school enjoyment. In terms of mechanisms, we find support for both a child and a parental effect. As hypothesized, treated children are more likely to receive high track recommendations and parents of treated children turn out to overrule low track recommendations more often than parents without treatment. The latter finding is particularly interesting as it shows that mentoring can affect household decision-making above and beyond immediate effects on the child. From a political perspective, these effects are promising in particular as the program seems easily scalable and highly cost effective.

We make several contributions to the literature. First, by showing that educational segregation tends to reinforce socioeconomic inequalities, this study speaks to work on the implications of educational sorting. Proponents of tracking point towards efficiency gains in the educational process due to targeted instruction or peer effects (Betts, 2011) with potential positive learning effects in particular for high ability students (Duflo et al., 2011; Card and Giuliano, 2016; Fu and Mehta, 2018).[5] At the same time, as confirmed in our paper, early tracking can lead to mismatch and inequalities of opportunity. While some studies find no adverse long-term consequences of coincidental tracking decisions for individuals at the margin

[5]Most studies that find positive effects of tracking on high ability students focus on special programs for gifted children within a given type of school. Our context instead is one where children are tracked into programs that are predominantly located at different types of schools with different curricular and school leaving certificates. See Betts (2011) for a review on within school tracking.

(Dustmann et al., 2017), others conclude that early tracking tends to reduce educational outcomes on average and for low SES students in particular (Hanushek and Wössmann, 2006; Pekkala Kerr et al., 2013; Piopiunik, 2014; Schneeweis and Zweimüller, 2014). Moreover, tracking can contribute to reduced social mobility, i.e., by increasing the importance of parental background for educational or labor market outcomes (Dustmann, 2004; Bauer and Riphahn, 2006; Brunello and Checchi, 2007; Pekkarinen et al., 2009). Later tracking or a larger fraction of seats in academic track programs on the other hand tend to improve educational outcomes (Malamud and Pop-Eleches, 2011; Guyon et al., 2012). The aim of this paper is not to derive conclusions about structure of the education system as such. Instead, similar to Carlana et al. (2017), we show that early tracking can lead to mismatch that can reinforce inequalities in a potentially unfair way, although our focus is not on high achieving migrant students, but on low SES children in general.

Second, by showing that a mentoring intervention can improve track outcomes and reduce inequality of opportunity, we contribute to a literature on interventions or mentoring that aim to reduce socio-economic disparities in education. So far, large-scale and long-term evaluations of RCTs targeted at elementary school children are rare, mostly focused on in-school programs, and their evidence is mixed (for an overview and discussion, see Kautz et al., 2014). In general, these programs seem most effective in improving academic outcomes if they target social skills or better relationships between children and parents (Hawkins et al., 2008; Sorrenti et al., 2020). A wider literature exists on mentoring and in particular the Big Brothers Big Sisters program (Foster, 2001; Eby et al., 2008; Moodie and Fisher, 2009; Raposa et al., 2019). However, few of these studies conduct randomized evaluations. Moreover, they focus almost exclusively on adolescent youths (see, e.g. Grossman and Tierney, 1998). While some of these programs show positive effects in terms of functioning in school or reduced dropout rates (Hawkins et al., 2005; Goux et al., 2017) others display zero or adverse long-run effects (McCord, 1978; DuBois et al., 2002; Rodriguez-Planas, 2012*a*). We confirm these findings to the extent that we do not observe substantial program effects on academic performance in terms of GPA. However, because the program sets in just before a critical decision-period, its effects on child behaviors and parental decision-making are large enough for the program to have long-run ramifications on children's educational careers, reducing the inequality of opportunity divide.[6] In this sense, our results also relate to the

[6]Examples of other programs that focus on critical decision-points during educational careers are Hoxby and Turner (2015) and Bettinger et al. (2012).

discussion on social mobility and the intergenerational persistence of life outcomes (Case et al., 2002; Currie and Moretti, 2003; Aizer and Currie, 2014).

Last, our work relates to a broad literature on programs and RCTs that can help improve the lives of disadvantaged children. A prominent example are (early) childhood education or intervention programs (e.g., Currie, 2001; Heckman et al., 2010; Fryer et al., 2015; Alan and Ertac, 2018; Attanasio et al., 2020; Cappelen et al., 2020). Moreover, several evaluations have shown the effectiveness of initiatives that explicitly aim at helping poor children gain opportunities (Chetty et al., 2016; Oreopoulos et al., 2017). Finally, there exist several programs that focus on the reduction of crime or other risky behaviors, some of which also display gains in schooling outcomes (see, e.g., Heller et al., 2017).

The remainder of the paper is organized as follows. In the next section, we discuss the institutional background, recruitment, our measures, details of the mentoring intervention, and its implementation. Section 3 first presents descriptive evidence on prevailing SES gaps in high track attendance and deduces our main research question regarding the program's effect on children's schooling trajectories and SES-related inequalities of opportunity. Subsequently, it unveils the causal effect of mentoring on high track attendance and provides suggestive evidence for underlying mechanisms. Finally, Section 4 concludes.

2 Institutional Background and Data

This section describes the German education system, the briq family panel data, and the concept of the mentoring intervention. We start out by explaining the German school system and the procedures of early tracking. Then, in section 2.2, we provide details on the sample and data collection, how we classified families in terms of SES, and how children were randomly assigned to three groups: control low SES, treatment low SES, and control high SES. Section 2.3 describes all relevant measures and variables and section 2.4 describes the mentoring intervention.

2.1 Institutional background

Children enter school at age six and then stay for four years of general elementary school education.[7] These elementary schools tend to be homogeneous in quality and are predominantly public or publicly financed (more than 90% of students attend

[7]Exact rules may slightly vary across time and federal state. We describe the system as it applies to children in our sample.

schools that are free of charge). Education there takes place in classes of around 21 children (Baumann et al., 2012) with a single head teacher who teaches all main subjects (German, Math, English).

After grade four, usually at the age of ten, students transition from primary school to a secondary school program. These programs differ in (academic versus vocational) content, length (5 to 9 years), and type of degree certificate that can be obtained. We classify these programs into "high track" and "low track" programs whereby only the high track leads to an upper secondary school degree, one that qualifies them for higher education studies. The high track lasts mostly nine years and the type of education prepares students for tertiary studies at academic institutions such as four-year colleges or five-year universities. The low track lasts five to six years and the school content is mostly vocational, i.e., preparing students for an apprenticeship.[8]

Track choice determines the type of classroom experience a child receives during grades five to ten (less than 2% of students switch tracks during that time, Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2017*b*). Overall educational expenditures are five percent higher in the high track and teachers receive 13-14% higher gross pay (Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2017*a*; German Statistical Office, 2016). Moreover, high track students are taught around 10% more hours and the breadth and intensity of academic subjects is substantially larger. They are also surrounded by better performing peers, as performance of high versus low track students in the Programme for International Student Assessment (PISA) test differs by more than one standard deviation (see, Dustmann et al. (2017) for a more detailed discussion of track differences).

There is no strict grade cutoff or central examination to determine which track children attend after elementary school. Instead, the institutional setting stipulates a two-stage decision process that involves both teachers and parents. In a first instance, all children receive a written teacher track recommendation on their midterm transcript of grade four. This recommendation is based on the evaluation of a child's performance and abilities by the student's head teacher and by other teachers, if applicable. The teacher recommendation is thus a clear and transparent evaluation of a child's abilities and an important signal to the parents. Then, in a second instance, usually in February/March of grade 4, parents select secondary schools based on the teacher recommendation and their own assessment, aspirations, and preferences (see figure A3 for a timeline of the decision-process). Importantly,

[8] Appendix B.1 describes the classification of school programs.

in this second step, parents can *choose freely* between both types of secondary school programs and schools are not allowed to select students based on ability or prior performance.[9] While the education system allows for later track-reversals (see, e.g., Tamm (2008); Dustmann et al. (2017)), initial track choices tend to be highly predictive of final education outcomes in particular among low SES children (Biewen and Tapalaga, 2017; Glaesser and Cooper, 2011).

2.2 Sample

The briq family panel comprises children born between 09/2002 and 08/2004 from the regions of Bonn and Cologne (Germany) who were contacted using official registry data.[10] To enter our study all families had to respond to an initial contact letter by returning a questionnaire about the socio-economic characteristics of the household. They also had to express interest in taking part in the mentoring program and accompanying interviews. For a more detailed description of the recruiting procedure see Appendix B.2 and Falk and Kosse (2020).

Based on the initial screening questionnaire, families were classified as high or low SES. Low SES families meet at least one of the following three criteria. First, *low income* describes families whose equivalence household income is lower than the 30th percentile of the German income distribution (1,065 Euro monthly net household equivalence income).[11] Second, in *low education* families neither the mother nor the father of the child have obtained a high track school degree, i.e., which qualifies for university studies. Third, *single parent* families are households where the child's primary caregiver is not living together with a partner. Households for which none of the three criteria apply are classified as high SES.

Low SES households were the target group of the mentoring program and our randomized control trial. We therefore invited all low SES families to take part in

[9] This rule of free ability-independent parental track choice is written into state law, and strictly enforced. Schools may give priority on the basis of the following criteria only: siblings, gender ratio, balancing of native and non-native speakers, distance to school, lottery. In practice, most schools give priority to siblings and then use a lottery to choose among all other applicants (Ministerium and des Landes Nordrhein-Westfalen für Schule und Bildung, 2017).

[10] Invitations to take part in the mentoring program and the interviews were sent to all families with children born between 09/2003 and 08/2004 and to one-third of the families with children born between 09/2002 and 08/2003. Birth dates were chosen such that children of the younger cohort were typically in second grade.

[11] As a reference, we make use of data from the German Socio-Economic Panel (SOEP, 2011), a representative sample of the German population consisting of more than 20,000 individuals in more than 10,000 households. For details see (Wagner et al., 2007). Equivalence income is computed according to the procedure used by the OECD and EUROSTAT (see Hagenaars et al., 1994). The resulting poverty threshold is (approximately) equivalent to the official poverty threshold, which in Germany is computed relative to the median income (e.g., 1,033 Euro in the year 2015).

the interviews. To be eligible for treatment, they had to participate in the baseline interviews and experiments (Fall 2011) and to provide written consent to allow the transmission of their addresses to the organization running the mentoring program (Balu und Du e.V.). Written consent could be given during the interview or be sent by postal mail thereafter. Out of 590 eligible families, 212 were randomly selected and form the intention-to-treat (ITT) group (Treatment Low SES). The remaining 378 form the control group (Control Low SES). We used stratified random treatment assignment to ensure a proportional representation of all forms of pre-defined disadvantage (low income, low education, single parenthood) in the ITT group, while making sure that the number of selected children matched the local supply of mentors.[12] After realization of the baseline interviews, contact information of the treatment group children was transmitted to the organization running the program (Balu und Du e.V.). As a second control group we also invited (in the same period) 150 randomly chosen high SES families (among those who had answered the information letter) to take part in the study (Control High SES). To ensure comparability we also asked them to give written consent to allow the transmission of their addresses to the organization running the mentoring program. 122 families took part at baseline and gave written consent. All families who had participated in the pre-treatment data collection were invited to take part in the post-treatment data collection and yearly follow-up data collections, for details see Appendix B.2.

Our working sample consists of all children for whom we have information about track attendance in grade 5. We also require information about track choice and a basic set of control variables (see section 2.3) to be non-missing. This leaves us with a final sample of 495 children.[13] At baseline, children where on average 7.8 years old and attended grade 2 or 3. In the analyses, we use data from parents and children collected between the end of 2011 (baseline) and the end of 2018 (6-year follow-up) when children were in grade 9 or 10. At each data collection point, parents and children were interviewed by trained interviewers who followed a rigorous protocol (see Appendix B.2 for details).

[12]The combination of the three forms of disadvantage and two regions yields 14 strata. Given the larger relative supply of mentors in Bonn, we also assigned a higher share of children in Bonn to the ITT group. Therefore, assignment into treatment was random conditional on region of residence.

[13]Attrition across data collections is non-selective in a sense that attrition is neither different across treatment groups ($p = 0.765$, two-sample test of proportions) nor does it relate to the interaction of baseline GPA and treatment status (see table A2 and section 3.3 for a discussion).

2.3 Variables and measures

In each interview phase, the children's mothers completed an extensive questionnaire covering their socio-economic background, information about school choice, as well as child school performance.

2.3.1 School choice

School type From the mother questionnaires, we derive information about the type of secondary school a child is attending. Education programs which lead to an upper secondary school degree that qualify for university studies (German *Abitur*) are classified as "high track". All other programs are classified as "low track". The German school system and the degree to which high and low track programs differ in quality are described in section 2.1.[14]

Teacher recommendation At the end of primary school, all parents receive a teacher recommendation regarding the most suitable track type for their child.[15] See section 2.1 for a description of the track choice process. The recommendation was elicited using the mother questionnaires in 2014. We generate a dummy variable which equals one if a child received a recommendation for the high track, and zero otherwise.

2.3.2 Child school performance

Elementary school GPA Grades at the end of the school year in German and mathematics were reported by the mothers. Since children are in secondary schools, we also elicited the child's grade in the first foreign language (usually English). In the German grading system, grades range from one to six where lower grades indicate better performance. Grade point average (GPA) is the equally weighted score of grades in German and mathematics when the children are still in elementary school. From secondary school onward, GPA also comprises the final grade in the first foreign language. For ease of interpretation we transform GPA, such that higher values indicate better performance. In the analysis we use the variable "elementary school GPA" as control variable. It is calculated as the average of the GPAs in the respective years of elementary school.

[14] See Appendix B.1 for details on the German secondary school system and an exact description of the tracking classification.

[15] Children have one head teacher during elementary school (teaching all main subjects) and this teacher usually consults with other teachers (e.g., teaching religion or sports) and the headmaster before giving a recommendation.

2.4 The mentoring intervention

The intervention that we implemented and evaluate is a well-established non-profit mentoring program called "Balu und Du" (German for "Baloo and you"). In this program, elementary school children are provided with a mentor for the duration of one year. The mentors, called Baloos, are university students (age 18 to 30) who typically spend one afternoon per week one-on-one with their mentees, called Mowglis. The conceptual idea of the program is for the mentors to act as friends and "big buddies" who encourage the acquisition of new ideas and skills by enriching the social environment of the children.[16] In addition to being pedagogically good role models, mentors tend to be more educated than the children's regular caregivers. In our sample, all mentors have obtained a high track degree and are currently enrolled at university. Thus, they may act as motivators and informants, e.g., about the schooling system and benefits of high track attendance.[17]

The mentoring program "Baloo and you" is embedded in a professional structure. On a weekly basis, mentors complete an online diary in which they report the activities that they have engaged in, as well as potential problems of the mentor-child relationship. Program coordinators read and comment on these diaries, and provide support. These coordinators are trained and paid professionals in education science or psychology and they provide supervision and advice to mentors. They also organize bi-weekly monitoring meetings where mentors receive suggestions for activities with the mentored child and discuss potential problems.[18]

We transmitted household addresses of all randomly-selected families to the mentoring organization. The actual matching process of mentor and mentee is part of the program and was conducted by the organization. Each child in the ITT group could potentially be matched, but not all selected children were effectively matched with a mentor. A mentor-mentee match was successfully implemented for 72% of the ITT group children. For 28% of the children, matches could not be realized due to a local shortage of mentors, mentor refusals or coordination problems between

[16] For further details, see Müller-Kohlenberg and Drexler (2013) and Kosse et al. (2020).

[17] The program is similar to many existing programs in the US and abroad. Examples are Big Brother/Big Sister, College Mentors for Kids, Friends of the Children, Mentoring USA and SHINE mentoring (Grossman and Tierney, 1998; Moodie and Fisher, 2009). However, most of these other programs tend to focus on teenagers (see Foster, 2001; Rodríguez-Planas, 2012*b*; Kautz et al., 2014, for reviews).

[18] The program has been honored with several public awards, e.g., by the Robert Bosch Foundation in 2011 and the federal government of North Rhine-Westphalia (Germany) in 2006. More details about the mentoring program can be found on www.balu-und-du.de and in an overview article by Müller-Kohlenberg and Drexler (2013).

mentors and families (e.g., pregnancy of the mentor, moving of mentor or family). Most of these children were never contacted by the organization.

3 Results

In this section, we first provide descriptive evidence on the (in)equality of opportunity divide, prevailing as Control High SES children are more likely to attend the high track in the year after tracking (grade five) than similarly performing Control Low SES children. Section 3.2 then describes the causal short and long-term effects of mentoring on the schooling decision. Here we also provide suggestive evidence on the program's effect in terms of child engagement at school and parental decision-making, respectively. Last, we provide the results from a substantial number of robustness analyses. This includes tests for baseline balance and the absence of selective attrition, an assessment of the external validity of our results, and local average treatment effect estimates.

For most analyses, we use data from the 2-year follow-up data collection, corresponding to the point in time when all children have transitioned to secondary school ($N = 495$). However, to study long-run effects of the intervention on educational trajectories, we extend our analysis up to the most recent wave of interviews (6-year follow-up) when children are in grade 9 or 10, i.e., five years after tracking ($N = 479$).

3.1 The (in)equality of opportunity divide

Research suggests a strong association between parental socio-economic status and the first schooling transition in Germany (e.g., Dustmann, 2004; Bauer and Riphahn, 2006; Biewen and Tapalaga, 2017). This is also apparent in our data. The estimates in column 1 of table 1 stem from a linear probability model of high track attendance on a dummy variable for high SES. They reveal that children from high SES backgrounds are 31.1 percentage points, or 63.7 percent, more likely to attend the high track ($N = 350$) after the schooling transition ($p < 0.01$).

While this raw gap is a meaningful description of socioeconomic inequalities in transition probabilities, it does not necessarily depict unfair access to educational opportunities. Instead, high SES children might simply be rewarded for working harder. They might also be more skillful, in part, because they benefit from better initial endowments or parental inputs, due to the "accident of birth" (Smith et al.,

1997; Sirin, 2005; Cunha and Heckman, 2007; Heckman, 2008).[19] Such differences in effort or skill naturally affect school performance. In our sample, they result in 40 percent of a standard deviation higher GPA among high SES children (see table A3).[20] Moreover, a one standard deviation increase in GPA then translates into a 24 percent higher probability of attending the high track during the year after the transition (table A4).

	High track attendance					
	in 5th grade			in 9/10th grade		
	(1)	(2)	(3)	(4)	(5)	(6)
Mean low SES:		0.488			0.508	
High SES dummy	0.311***	0.309***	0.217***	0.373***	0.374***	0.290***
	(0.054)	(0.054)	(0.051)	(0.048)	(0.049)	(0.049)
Controls:						
Age & sex	No	Yes	Yes	No	Yes	Yes
GPA (elementary)	No	No	Yes	No	No	Yes
Observations	350	350	350	342	342	342
R-squared	0.076	0.088	0.303	0.115	0.128	0.271

Table 1: SES gaps at the point of tracking and at the end of secondary schooling. Coefficients are OLS estimates using location FEs with White robust standard errors in parentheses. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

In addition to looking at raw SES gaps in transition probabilities, we thus also investigate gaps net of school performance differences. By eliminating differences due to (righteous) effort and (potentially tilted) skill levels, they can be seen as a lower bound measure of prevailing unfairness in the access to academic education.[21] The results displayed in columns 2 and 3 of table 1 reveal that age and sex hardly affect the estimated high SES coefficient (column 2), while elementary school GPA reduces it (column 3). Yet, performance differences account for only *less than one third* of the gap in high track attendance between Control High SES and Control Low SES children. This finding is surprising and reveals a high level of inequality of opportunity. It indicates that the educational career of a child seems to be more

[19] See also Schwippert et al. (2003) for evidence on skill differences in Germany.

[20] Assuming that grading is non-discriminatory against low SES children. See Alesina et al. (2018) for evidence on discrimination against immigrants.

[21] This relates to a larger literature in sociology on primary (performance related) and secondary (non-performance related) effects in educational decision-making (see, e.g., Jackson et al., 2007; Kloosterman et al., 2009; Neugebauer, 2010) going back to Boudon (1974).

strongly determined by her parents' socioeconomic resources than by her own efforts and abilities.[22]

Do the above-described unconditional and conditional SES gaps have long-run ramifications? When looking at educational trajectories until adolescence, we find that both gaps increase over time. At the end of secondary school (grades nine/ten) high SES children are overall 37.3 percentage points (versus 31.1 pp in grade five) more likely to attend a high track program and 29 percentage points (versus 21.7 pp in grade five) after accounting for elementary school performance (see columns 4 and 6 of table 1). This increase in socioeconomic inequality is in line with prior results by, e.g., Biewen and Tapalaga (2017) who show that low SES children are much less likely to transition to a high track program at later stages of secondary school than high SES children. It suggests that unequal opportunities at early stages of the educational career are enduring and particularly detrimental for low SES children.[23]

While socio-economic inequalities in access to educational opportunities are particularly salient in a situation characterized by early track choices and transparent school differences, they are symptomatic of similar opportunity gaps across the world (Corak, 2013; Carlana et al., 2017; Walters, 2018; Rothstein, 2019). As regards their sources, a lack of access to high SES role models has been identified as an important determinant of "opportunity gaps" faced by low SES children (Chetty et al., 2016; Putnam, 2016). Favorable role models have the potential to increase a child's educational motivation and engagement (Rhodes et al., 2000) and can inform or convince her parents about the feasibility and benefits of education (Cunha et al., 2013; Boneva and Rauh, 2018). Following this line of thought, our hypothesis is that access to a high SES mentor might increase the transition probability of low SES children and reduce part of the above-described inequality of opportunity divide.

In the remainder of the paper, we thus address the following research question: Can access to a highly educated mentor causally reduce the unconditional and conditional SES gaps in access to education? Our aim is to unveil the causal effect of

[22]Our sampling scheme allows us to take a closer look at transition probabilities within groups of low SES children that differ by type of socio-economic disadvantage (low income, low education, single parent). Because these groups are non-exclusive and many children face multiple forms of disadvantage, we estimate a model with three indicators as explanatory variables. Findings indicate that the inequality of opportunity divide is almost exclusively driven by children who grow up in low income and low parental education families (see table A5), while children from single parent families are hardly disadvantaged.

[23]While startling, these gaps are no larger in our sample than in the general population of children. Using representative SOEP-data for 17 year-old children, we find an SES gap of 35.1 percentage points ($p < 0.01$). Moreover, these representative findings mirror our data as differences are almost exclusively driven by children from low-income and low parental education families (see table A6).

mentoring on children's educational trajectories both in the short run and in the longer run, and to provide suggestive evidence on the effect of mentoring on children and parents, respectively.

3.2 The causal effect of mentoring on schooling decisions

Panel A of figure 1 displays the effect of the mentoring program on tracking outcomes among Treatment Low SES children and the extent to which these children catch up with the High SES Control group. Right after the transition (grade five), the Treatment Low SES children are 11 percentage points (21.7 percent) more likely to attend the high track than those from the Low SES control group ($p < 0.05$, see table A7 column 1). The treatment thus closes about one third of the unconditional 31 percentage point gap between the Control Low SES and the Control High SES groups (see also section 3.1).[24] In terms of magnitude this effect is sizable; it is about twice as large as the effect of being roughly one year older at school entry (i.e., of being born just before as opposed to just after the school entry age cutoff, see Dustmann et al., 2017) and nearly as large as the associative impact of a 0.5 sd increase in GPA (see table A4).

Does the interaction with a highly educated mentor also reduce the performance-adjusted SES gap? Results displayed in Panel B of table 1 suggest that this is indeed the case. When comparing the bars in the top part of panel B, two main findings emerge. First, the estimated effect size of the mentoring program remains the same after controlling for basic demographic characteristics and elementary school GPA (0.110, $p < 0.05$), while its precision increases. This finding confirms that treatment assignment within the low SES group of children was independent of school performance. Second, participation in the mentoring program closes roughly *one half* of performance-adjusted SES gap, reducing it from 21.7 percentage points to 11.2 percentage points.

[24] See section 3.3 for robustness analyses.

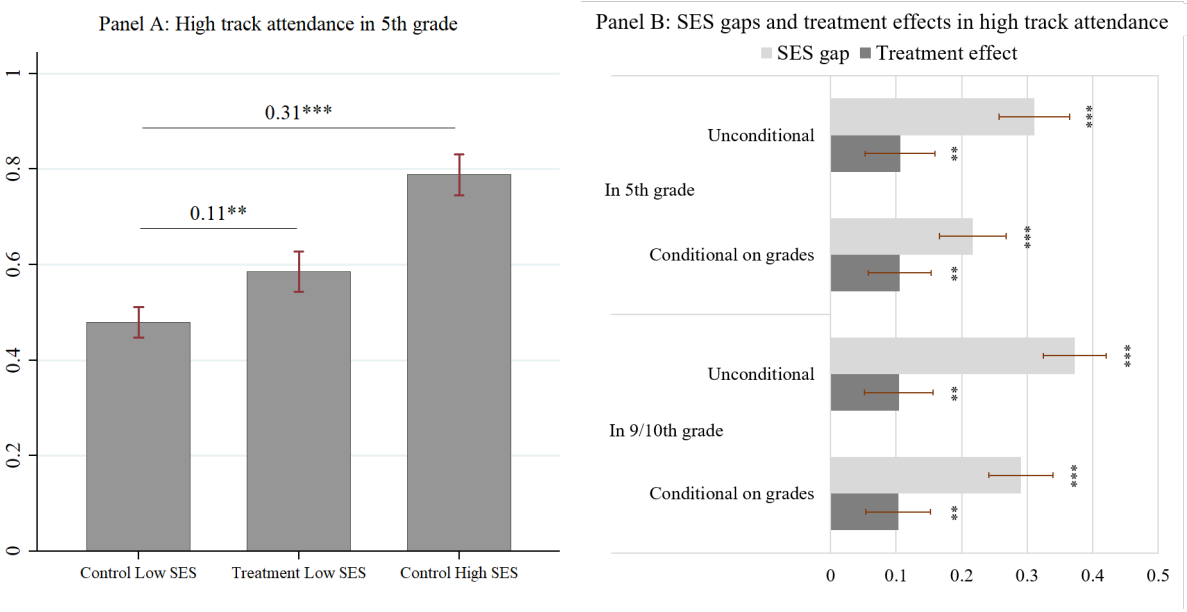

Figure 1: **Panel A** displays the fraction of children who attend the high track in 5th grade by treatment assignment category. Error bars reflect standard errors. Results correspond to a linear probability model with location FEs. **Panel B** displays unconditional and conditional SES gaps and treatment effects (ITT) in high track attendance in grades 5 and 9/10. Unconditional gaps are estimated differences without further controls. Conditional gaps are estimated differences conditional on GPA in elementary school, sex and age. Details regarding the SES gap in grade 5 are shown in table 1 (columns 1 and 3), details regarding the treatment effect in 5th grade are shown in table A7 (columns 1 and 3), details regarding the SES gap in 9/10th grade are shown in table 1 (columns 4 and 6), and details regarding the treatment effect 5 years after tracking are shown in table A8 (columns 1 and 3). ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

We now turn to the important question of whether mentoring can improve children's educational trajectories and well-being also in the longer run. Anecdotal evidence suggests that the most important rationale for sending low SES children to a low track program among both teachers and parents is the concern that difficult material and high work load might lead to distress in the medium and longer term. One worry is that children with lower educational resources at home might have to repeat a grade or be "down-tracked" after a while, leaving them with a sense of failure. This is, however, not what we find. Instead, the results displayed in the bottom part of panel B in figure 1 suggest that the treatment effect is enduring and of about equal size (point estimate of 0.104, $p < 0.05$) even five years after tracking. Figure A4 indicates that not only the treatment effect is persistent over time, but also the levels of high track attendance in treatment and control groups. Moreover, the treatment effect remains almost exactly the same (0.103, $p < 0.05$) in terms of

size and significance even after we control for age, sex and elementary school grades (see table A8). Overall, the program closes more than one-third of the long-run performance-adjusted SES gap (see bars at the bottom of panel B).

We proceed by investigating the dynamics of track attendance between the 5th and 9/10th grade to evaluate whether this long-term effect is indeed driven by the *same* children who benefited from the mentoring program in terms of their school track choice after grade four. Figure A5 displays the probability of being tracked up, tracked down, and of having to repeat a grade in the treatment and control groups, respectively. The figure illustrates two important findings. First, it is relatively uncommon that initial track choices are reversed between grades 5 and 9/10 (see the left and middle panels). As a consequence, the initial decision after grade four is highly predictive of the type of education low SES children receive during the ages 10-16.[25] Second, track reversals and grade retention are equally likely in the treatment and control groups, suggesting that those children who attend the high track *because of* the mentoring program are no more likely to experience failure. Arguably, these findings do not exclude the possibility that children in the treatment group experience higher levels of stress, e.g., due to lower grades. However, we again find no evidence of this presumption. The effects on GPA are small and not significantly different from zero (see the results in table A9). Moreover, there are no adverse effects in terms of school enjoyment: differences in school satisfaction ("I enjoy going to school") between children in the Low SES Treatment and Control groups are small and non-significant (see table A10). However, overall, children in high track programs display 30 percent of a standard deviation higher school enjoyment than those in low track programs.

In the following, we zoom into the decision-making process to obtain a better understanding of the program's impacts. Decision-making about track attendance involves three types of actors: The child, whose academic potential is being evaluated; the head teacher, who has known the child for several years and who gives a track recommendation assessing her long-run academic potential; and the parents, who are the legal decision-makers. As mentoring takes place in the household context, it only affects children and parents.[26] The effect on the child is likely to manifest itself mainly through a better teacher recommendation, which above and beyond grades, reflects an assessment of the child's potential and motivation to succeed in a high track program. The effect on the parents is likely to manifest itself in

[25] See Biewen and Tapalaga (2017); Glaesser and Cooper (2011) for prior evidence showing that initial tracking strongly determines final education outcomes.

[26] As a rule, teachers were not informed about the program, but the possibility exists that children talked about their mentors at school.

their decision-making after having received the teacher's recommendation. To provide suggestive evidence on the relative effect of the program on children and their parents, we provide a simple nonparametric decomposition of the overall treatment effect, dissecting it into program-induced changes in teacher recommendations and parental decision-making (see Appendix C for details). We find that roughly one third of the overall effect can be ascribed to the positive impact of the mentors on the children and the resulting changes in teacher recommendations. Approximately two-thirds of the program effect are due to changes in parental decision-making (see figure C1). Thus, the program seems to encourage low SES children to strive at school, which results in better teacher recommendations. However, it also seems to induce low SES parents to behave more like high SES parents in that they increasingly follow a high teacher recommendation and are more inclined to overrule a low one. As both effects contribute to the reported SES gaps (see Appendix D on SES differences in teacher recommendations and parental decision-making), and are mitigated by the treatment, this brings our findings to full circle.

3.3 Robustness checks and local average treatment estimates

The aim of this section is to assess the robustness and external validity of our results, and to present alternative treatment estimates. We start by investigating baseline imbalance. Then, to assess systematic attrition, we discuss results based on weights or treatment-effect bounds that account for the probability of being lost to follow-up. Additionally, we present evidence based on representative population weights to appraise the external validity of our results. Last, we discuss compliance and present local average treatment effects using random treatment assignment as an instrument for program take-up.

A potential concern for the validity of our estimates is whether the randomization procedure indeed yielded observationally equivalent treatment and control groups. In this respect, an important feature of our study design is that we have also collected measures of academic performance (school grades) *before* treatment assignment. Using these data, we can show that the randomization procedure was successful: pre-treatment GPA is not associated with treatment status (coefficient of -0.002, see column 1 of table A2).

A second potential limitation for the interpretation of randomized intervention studies is selective attrition. In our case, one may worry that study participation in later data collections systematically relates to treatment status, which would bias our estimates. Moreover, treatment status might differentially affect attrition among

high or low performing students. In columns 2 and 3 of table A2, we provide estimates of the probability to being lost to follow-up as a function of treatment status and GPA. The results indicate that attrition in grade 5 and 9/10, respectively, are not related to treatment assignment nor to its interaction with GPA. The treatment effects thus prove remarkably robust to potential selective attrition. In fact, the results hardly change when we re-weigh the observed data using inverse probabilities of participation in grade 5 and 9/10 (see table A11, columns 2 and 5, and Appendix B.3.1 for details). Nevertheless, we further assess bias due to selective attrition by estimating treatment effects using the trimming procedure suggested by Lee (2009). Instead of correcting point estimates, this approach yields interval estimates of effect sizes on the basis of extreme assumptions about selection. Given the near absence of systematic attrition discussed above (compare table A2), it is not surprising that the analysis yields tight bounds in line with previously reported results. The Lee-bounds of the treatment effects are 0.104 to 0.112 for high track attendance in grade 5 and 0.088 to 0.129 for high track attendance in grade 9/10, respectively (see table A12).

The above suggests that our findings are robust and internally valid, but do they generalize? Arguably, although the initial recruitment was based on official registry data, families in our sample are not necessarily representative of the German population. After all, any family taking part is low SES (in at least one dimension) and declared interest in the mentoring program. Consequently, the reported effect sizes constitute average effects of voluntary mentoring for low SES children. Other types of policy-relevant effects are conceivable, however. For example, a universal roll-out of the program might induce a much larger, representative pool of families to participate. To assess the program's external validity and its scalability to a wider group of families, we construct representative population weights that account for the probability of being part of our sample as a function of the three SES dimensions, child elementary school GPA and its interaction (see Appendix B.3.2). This is possible because we collected this information corresponding to the questions posed to SOEP participants, a data set that is representative of the German population at large. The reweighed ITT effects, displayed in columns 3 and 6 of table A11, are very similar to the main effects displayed in columns 1 and 4. This provides tentative evidence that ITT effects might be similar if the program was rolled out to a (more) representative group of families.

So far, all presented program effects result from a comparison of education outcomes between the respective treatment assignment groups. However, not all children who were initially assigned to the Low SES Treatment group have actually

participated in the mentoring program. For 28% of the treatment children in our sample, the mentor-mentee match was either not initiated due to a shortage of mentors or matches could not be realized due to moving, other coordination problems, or refusals by mentors or families.[27] The above-displayed effect is thus the intention-to-treat effect, i.e., the average effect of informal mentoring during elementary school if offered to interested low SES families on a voluntary basis. However, since the mentoring program takes effect through successful participation, we provide the corresponding local average treatment effects (LATE), that is, the average effect among the treatment compliers.[28] Table A13 displays two-stage least squares (2SLS) estimates that use the random assignment as an instrument for treatment receipt. In line with a matching rate (compliance rate) of around 72%, the LATE amounts to 15 percentage points immediately after the transition and to 14 percentage points in the long run, i.e., exceeding the ITT effect by about 38%. In terms of the equality of opportunity divide, the causal local average treatment effect of the program thus closes more than two-thirds of the short-run and more than half of the long-run conditional SES gap.

4 Discussion and Conclusion

This paper has reported good and bad news. The bad news is that access to high quality education not only depends on merit but also on a child's socio-economic background. Conditional on school performance prior to selection, children from high SES background are significantly more likely to attend the high track than their low SES counterparts. Given the importance of tracking for labor market and many other outcomes later in life, this unfair selection reflects a substantial lack of inequality of opportunity. What's more, a system that discriminates along parental background rather than merit is likely to be economically inefficient. The reason is that many children's potential will not be realized, and the benefits arising from complementarities between skills and high quality education are not harnessed.

At the same time, childhood intervention programs can help level the playing field. This is the good news the paper reports on. Our main result shows that social programs, such as mentoring, can substantially raise disadvantaged children's

[27] In terms of observable characteristics, complier and non-complier children do not significantly differ in any relevant dimension (see also Kosse et al., 2020, for a discussion).

[28] Among complier children the average number of meetings was 22.6 (sd 11.8) and the treatment duration was 9.9 months (sd 4.3). Hence, there is considerable variation in treatment intensity. This is however highly endogenous as children with more difficulties tend to see their mentors more often and for a longer period of time.

educational prospects. We find that the SES gap in high track attendance that arises conditional on GPA, can be reduced by half. Children who receive support from a high-skilled mentor and through a program designed to foster informal learning and psycho-social development yield better assessments by teachers and the program also raises confidence of parents in their children. Both effects contribute to the higher likelihood of high track attendance among treated low SES students relative to the control group. Importantly, these positive effects are lasting, and our data do not suggest that treated children suffer from increased grade retention, worse grades or less school enjoyment.

The nature of the discussed program has two features noteworthy from a policy perspective. First, it is fairly cost effective. The cost of a mentor-mentee pair amounts to roughly 1,000 Euro in addition to a mentor's opportunity cost of time. This cost should be related to the expected economic benefits that arise from substantially higher labor market returns for children entering the high rather than the low track. Of course, selection and general equilibrium effects will complicate a precise estimate of costs and benefits. Even the most conservative estimate of the benefits arising from higher expected life-time earnings, however, will easily exceed the cost of the program. Moreover, mentoring disadvantaged children yields societal benefits above and beyond economic returns. For example, Kosse et al. (2020) show that treated children are significantly more prosocial than those in the control group. Similarly, children participating in the mentoring program are found to be less likely to lie than their control group counterparts (Abeler et al., 2020), a behavior clearly beneficial from a society's point of view.

Second, the value of a social program depends critically on its scalability. In this respect, the mentoring program under study qualifies well. As of today, Baloo and You has initiated 12,275 mentor-mentee pairs, is operating in 104 cities and is steadily increasing its activity. Moreover, it is plausible that the reported effects are at work in related programs that similarly foster relationships and role modeling (e.g. Big Brother/Big Sister, College Mentors for Kids, Mentoring USA, SHINE mentoring etc.). The low cost and informal nature of mentoring programs make it relatively easy to further expand their activities. In light of our findings, this would be for the better.

References

Abeler, J., Falk, A. and Kosse, F. (2020), Determinants of preferences for honesty. Mimeo.

Aizer, A. and Currie, J. (2014), 'The intergenerational transmission of inequality: Maternal disadvantage and health at birth', *Science* **344**(6186), 856–861.

Alan, S. and Ertac, S. (2018), 'Fostering patience in the classroom: Results from randomized educational intervention', *Journal of Political Economy* **126**(5), 1865–1911.

Alesina, A., Carlana, M., Ferrara, E. L. and Pinotti, P. (2018), Revealing stereotypes: Evidence from immigrants in schools. NBER Working Paper No. 25333.

Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C. and Rubio-Codina, M. (2020), 'Estimating the production function for human capital: Results from a randomized controlled trial in Colombia', *American Economic Review* **110**(1), 48–85.

Bauer, P. and Riphahn, R. T. (2006), 'Timing of school tracking as a determinant of intergenerational transmission of education', *Economics Letters* **91**(1), 90–97.

Baumann, T., Schneider, C., Vollmar, M. and Wolters, M. (2012), *Schulen auf einen Blick*, Wiesbaden: Statistisches Bundesamt.

Bettinger, E. P., Long, B. T., Oreopoulos, P. and Sanbonmatsu, L. (2012), 'The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment', *The Quarterly Journal of Economics* **127**(3), 1205–1242.

Betts, J. R. (2011), The economics of tracking in education, *in* 'Handbook of the Economics of Education', Vol. 3, Elsevier, pp. 341–381.

Biewen, M. and Tapalaga, M. (2017), 'Life-cycle educational choices in a system with early tracking and "second chance" options', *Economics of Education Review* **56**, 80–94.

Boneva, T. and Rauh, C. (2018), 'Parental beliefs about returns to educational investments — the later the better?', *Journal of the European Economic Association* **16**(6), 1669–1711.

Boudon, R. (1974), *Education, Opportunity, and Social Inequality: Changing Prospects in Western Society.*, New York: Wiley.

Brunello, G. and Checchi, D. (2007), 'Does school tracking affect equality of opportunity? New international evidence', *Economic Policy* **22**(52), 782–861.

Cappelen, A. W., List, J. A., Samek, A. S. and Tungodden, B. (2020), 'The effect of early education on social preferences', *Journal of Political Economy* (forthcoming).

Card, D. and Giuliano, L. (2016), 'Can tracking raise the test scores of high-ability minority students?', *American Economic Review* **106**(10), 2783–2816.

Carlana, M., La Ferrara, E. and Pinotti, P. (2017), Goals and gaps: Educational careers of immigrant children. CEPR Discussion Papers No. 12538.

Case, A., Lubotsky, D. and Paxson, C. (2002), 'Economic status and health in childhood: The origins of the gradient', *American Economic Review* **92(5)**, 1308–1334.

Chetty, R., Hendren, N. and Katz, L. F. (2016), 'The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment', *American Economic Review* **106**(4), 855–902.

Corak, M. (2013), 'Income inequality, equality of opportunity, and intergenerational mobility', *Journal of Economic Perspectives* **27**(3), 79–102.

Cunha, F., Elo, I. and Culhane, J. (2013), Eliciting maternal beliefs about the technology of skill formation. NBER Working Paper No. 19144.

Cunha, F. and Heckman, J. (2007), 'The economics of human development the technology of skill formation', *American Economic Review* **97**(2), 31–47.

Currie, J. (2001), 'Early childhood education programs', *Journal of Economic perspectives* **15**(2), 213–238.

Currie, J. and Moretti, E. (2003), 'Mother's education and the intergenerational transmission of human capital: Evidence from college openings', *The Quarterly Journal of Economics* **118**(4), 1495–1532.

DuBois, D. L., Holloway, B. E., Valentine, J. C. and Cooper, H. (2002), 'Effectiveness of mentoring programs for youth: A meta-analytic review', *American Journal of Community Psychology* **30**(2), 157–197.

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

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

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

Eby, L. T., Allen, T. D., Evans, S. C., Ng, T. and DuBois, D. L. (2008), 'Does mentoring matter? A multidisciplinary meta-analysis comparing mentored and non-mentored individuals', *Journal of Vocational Behavior* **72**(2), 254–267.

Falk, A. and Kosse, F. (2020), The briq family panel: An overview. Mimeo.

Foster, L. (2001), Effectiveness of mentor programs: Review of the literature from 1995 to 2000, Technical report, California State Library, Sacramento. California Research Bureau. Prepared at the Request of Senator Dede Alpert, Chair, Senate Select Committee On Family, Child and Youth Development.

Fryer, R. G., Levitt, S. D. and List, J. A. (2015), Parental incentives and early childhood achievement: A field experiment in Chicago heights. NBER Working Paper No. 21477.

Fu, C. and Mehta, N. (2018), 'Ability tracking, school and parental effort, and student achievement: A structural model and estimation', *Journal of Labor Economics* **36**(4), 923–979.

Garces, E., Thomas, D. and Currie, J. (2002), 'Longer-term effects of Head Start', *American Economic Review* **92**(4), 999–1012.

German Statistical Office (2016), Schulen auf einen Blick, Report, Statistisches Bundesamt.

Glaesser, J. and Cooper, B. (2011), 'Selectivity and flexibility in the German secondary school system: A configurational analysis of recent data from the German Socio-Economic Panel', *European Sociological Review* **27**(5), 570–585.

Goodman, J., Hurwitz, M., Mulhern, C. and Smith, J. (2019), O brother, where start thou? Sibling spillovers in college enrollment. NBER Working Paper No. 26502.

Goux, D., Gurgand, M. and Maurin, E. (2017), 'Adjusting your dreams? High school plans and dropout behaviour', *The Economic Journal* **127**(602), 1025–1046.

Grossman, J. B. and Tierney, J. P. (1998), 'Does mentoring work? An impact study of the Big Brothers Big Sisters program', *Evaluation Review* **22**(3), 403–426.

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

Hagenaars, A. J., De Vos, K. and Zaidi, A. (1994), Poverty Statistics in the late 1980s: Research based on Micro-Data, Technical report, Study carried out for Eurostat, Office for Official Publications of the European Community, Luxembourg.

Hanushek, E. and Wössmann, L. (2006), 'Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries', *The Economic Journal* **116**, 363–376.

Hawkins, J. D., Kosterman, R., Catalano, R. F., Hill, K. G. and Abbott, R. D. (2005), 'Promoting positive adult functioning through social development intervention in childhood: Long-term effects from the Seattle Social Development Project', *Archives of Pediatrics & Adolescent Medicine* **159**(1), 25–31.

Hawkins, J. D., Kosterman, R., Catalano, R. F., Hill, K. G. and Abbott, R. D. (2008), 'Effects of social development intervention in childhood 15 years later', *Archives of Pediatrics & Adolescent Medicine* **162**(12), 1133–1141.

Heckman, J. J. (2008), 'Schools, skills and synapses', *Economic Inquiry* **46**(3), 289–324.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A. and Yavitz, A. (2010), 'The rate of return to the HighScope Perry Preschool Program', *Journal of Public Economics* **94**(1-2), 114–128.

Heckman, J., Pinto, R. and Savelyev, P. (2013), 'Understanding the mechanisms through which an influential early childhood program boosted adult outcomes', *American Economic Review* **103**(6), 2052–86.

Heller, S. B., Shah, A. K., Guryan, J., Ludwig, J., Mullainathan, S. and Pollack, H. A. (2017), 'Thinking, fast and slow? Some field experiments to reduce crime and dropout in chicago', *The Quarterly Journal of Economics* **132**(1), 1–54.

Hoxby, C. M. and Turner, S. (2015), 'What high-achieving low-income students know about college', *American Economic Review* **105**(5), 514–17.

Jackson, M., Erikson, R., Goldthorpe, J. H. and Yaish, M. (2007), 'Primary and secondary effects in class differentials in educational attainment: The transition to A-level courses in england and wales', *Acta Sociologica* **50**(3), 211–229.

Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B. and Borghans, L. (2014), Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. NBER Working Paper No. 20749.

Kloosterman, R., Ruiter, S., De Graaf, P. M. and Kraaykamp, G. (2009), 'Parental education, children's performance and the transition to higher secondary education: Trends in primary and secondary effects over five Dutch school cohorts (1965–99)', *The British Journal of Sociology* **60**(2), 377–398.

Kosse, F., Deckers, T., Pinger, P., Schildberg-Hörisch, H. and Falk, A. (2020), 'The formation of prosociality: Causal evidence on the role of social environment', *Journal of Political Economy* **128**(2), 434–467.

Lee, D. S. (2009), 'Training, wages, and sample selection: Estimating sharp bounds on treatment effects', *The Review of Economic Studies* **76**(3), 1071–1102.

List, J. A., Samek, A. and Suskind, D. L. (2018), 'Combining behavioral economics and field experiments to reimagine early childhood education', *Behavioural Public Policy* **2**(1), 1–21.

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

McCord, J. (1978), 'A thirty-year follow-up of treatment effects.', *American Psychologist* **33**(3), 284–289.

Ministerium and des Landes Nordrhein-Westfalen für Schule und Bildung (2017), 'Verordnung über die Ausbildung und die Abschlussprüfungen in der Sekundarstufe I Ausbildungs-und Prüfungsordnung Sekundarstufe I-APO-S I, Beschluss vom 2. November 2012'.

Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (2017*a*), Besoldung bei Übernahme in den Schuldienst im Beamtenverhältnis, Report, Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen.

Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (2017*b*), Das Schulwesen in Nordrhein-Westfalen aus quantitativer Sicht 2016/2017, Statistische Übersicht 395, Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen.

Moodie, M. L. and Fisher, J. (2009), 'Are youth mentoring programs good value-for-money? An Evaluation of the Big Brothers Big Sisters Melbourne Program', *BMC Public Health* **9**(1), 41.

Müller-Kohlenberg, H. and Drexler, S. (2013), *Mentoring: Practices, Potential Challenges and Benefits*, Nova Science Publishers, New York.

Neugebauer, M. (2010), 'Bildungsungleichheit und Grundschulempfehlung beim Übergang auf das Gymnasium: Eine Dekomposition primärer und sekundärer Herkunftseffekte/Educational Inequality and Teacher Recommendations at the Transition to Upper Secondary School: A Decomposition of Primary and Secondary Effects of Social Origin', *Zeitschrift für Soziologie* **39**(3), 202–214.

Oreopoulos, P., Brown, R. S. and Lavecchia, A. M. (2017), 'Pathways to education: An integrated approach to helping at-risk high school students', *Journal of Political Economy* **125**(4), 947–984.

Pekkala Kerr, S., Pekkarinen, T. and Uusitalo, R. (2013), 'School tracking and development of cognitive skills', *Journal of Labor Economics* **31**(3), 577–602.

Pekkarinen, T., Uusitalo, R. and Kerr, S. (2009), 'School tracking and intergenerational income mobility: Evidence from the finnish comprehensive school reform', *Journal of Public Economics* **93**(7), 965–973.

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

Putnam, R. D. (2016), *Our kids: The American dream in crisis*, Simon and Schuster.

Raposa, E. B., Rhodes, J., Stams, G. J. J., Card, N., Burton, S., Schwartz, S., Sykes, L. A. Y., Kanchewa, S., Kupersmidt, J. and Hussain, S. (2019), 'The effects of youth mentoring programs: A meta-analysis of outcome studies', *Journal of Youth and Adolescence* **48**(3), 423–443.

Rhodes, J. E., Grossman, J. B. and Resch, N. L. (2000), 'Agents of change: Pathways through which mentoring relationships influence adolescents' academic adjustment', *Child Development* **71**(6), 1662–1671.

Rodriguez-Planas, N. (2012*a*), 'Longer-term impacts of mentoring, educational services, and learning incentives: Evidence from a randomized trial in the united states', *American Economic Journal: Applied Economics* **4**(4), 121–39.

Rodríguez-Planas, N. (2012*b*), 'Mentoring, educational services, and incentives to learn: What do we know about them?', *Evaluation and Program Planning* **35**(4), 481–490.

Rothstein, J. (2019), 'Inequality of educational opportunity? Schools as mediators of the intergenerational transmission of income', *Journal of Labor Economics* **37**(S1), S85–S123.

Schneeweis, N. and Zweimüller, M. (2014), 'Early tracking and the misfortune of being young', *The Scandinavian Journal of Economics* **116**(2), 394–428.

Schwippert, K., Bos, W. and Lankes, E.-M. (2003), 'Heterogenität und chancengleichheit am ende der vierten jahrgangsstufe im internationalen vergleich', *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* pp. 265–302.

Sirin, S. R. (2005), 'Socioeconomic status and academic achievement: A meta-analytic review of research', *Review of Educational Research* **75**(3), 417–453.

Smith, J. R., Brooks-Gunn, J. and Klebanov, P. K. (1997), *Consequences of Growing up Poor*, chapter Consequences of living in poverty for young children's cognitive and verbal ability and early school achievement.

SOEP (2011), 'German Socio-Economic Panel, Data for Years 1984-2009, Version 26', doi:10.5684/soep.v26. Dataset.

SOEP (2019), 'German Socio-Economic Panel, Data for Years 1984-2018, Version 35', doi:10.5684/soep-core.v35. Dataset.

Sorrenti, G., Zölitz, U., Ribeaud, D. and Eisner, M. (2020), The causal impact of socio-emotional skills training on educational success. Institute for the Study of Labor (IZA) Discussion Paper No. 13087.

Tamm, M. (2008), 'Does money buy higher schooling? evidence from secondary school track choice in Germany', *Economics of Education Review* **27**(5), 536–545.

Tauchmann, H. (2014), 'Lee (2009) treatment-effect bounds for nonrandom sample selection', *The Stata Journal* **14**(4), 884–894.

Wagner, G. G., Frick, J. R., Schupp, J. et al. (2007), 'The german socio-economic panel study (soep)–scope, evolution and enhancements', *Journal of Applied Social Science Studies* **127**(1), 139–169.

Walters, C. R. (2018), 'The demand for effective charter schools', *Journal of Political Economy* **126**(6), 2179–2223.

A Additional Tables and Figures

Tracking in OECD countries			
Country	Grade	Country	Grade
Austria	4	Israel	9
Germany	4	Japan	9
Czech Republic	5	South Korea	9
Ireland	6	Latvia	9
Luxembourg	6	Portugal	9
Mexico	6	Slovak Republic	9
Netherlands	6	Slovenia	9
Hungary	4/6/8	Spain	9
Belgium	8	Sweden	9
Italy	8	Switzerland	9
Lithuania	8	Australia	10
Poland	8	Chile	10
Turkey	8	Iceland	10
Denmark	9	New Zealand	10
Estonia	9	Norway	10
Finland	9	United Kingdom	10
France	9	Canada	No tracking
Greece	9	United States	No tracking

Table A1: The table indicates after which school grade tracking takes place in the respective country. Note: Quebec (Canada) tracks after grade 9.

	Assigned to treatment	Lost to follow-up	
		Grade 5	Grade 9/10
	(1)	(2)	(3)
Treatment dummy		-0.013	0.005
		(0.039)	(0.041)
GPA (baseline, std.)	-0.002	-0.046*	-0.043*
	(0.019)	(0.024)	(0.024)
GPA x Treatment Dummy		-0.043	-0.010
		(0.039)	(0.041)
Restricted to low SES	Yes	Yes	Yes
Observations	590	590	590
R-squared	0.000	0.020	0.011

Table A2: Checks for baseline balance and selective attrition. GPA is standardized and higher values indicate better grades. In column (1), we test for baseline balance. The dependent variable is one if a child was selected into the Treatment Low SES group and zero if selected into the Control Low SES group. In columns (2) and (3), we test for selective attrition. The dependent variable is one if a child is lost to follow-up, i.e., did not take part in the grade 5 or grade 9/10 interviews, and zero otherwise. Displayed coefficients are OLS estimates with White robust standard errors in parentheses. ***, **, * indicate significance at 1, 5 and 10 percent level, respectively.

	Elementary school GPA (standardized)	
	(1)	(2)
High SES dummy	0.390***	0.396***
	(0.109)	(0.109)
Controls:		
Age & sex	No	Yes
Observations	350	350
R-squared	0.037	0.045

Table A3: SES gaps in elementary school GPA. Coefficients are OLS estimates using location FEs with White robust standard errors in parentheses. Sample: low and high SES control groups. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

	High track attendance in 5th grade	
	(1)	(2)
Elementary school GPA (standardized)	0.247***	0.233***
	(0.017)	(0.018)
Controls:		
Age & sex	No	Yes
Observations	350	350
R-squared	0.252	0.293

Table A4: Elementary school GPA and high track attendance. Coefficients are OLS estimates using location FEs with White robust standard errors in parentheses. Sample: low and high SES control groups. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

	High track attendance in 5th grade		
	(1)	(2)	(3)
Low income	-0.183***	-0.181***	-0.134***
	(0.045)	(0.045)	(0.040)
Low education	-0.288***	-0.289***	-0.193***
	(0.045)	(0.045)	(0.043)
Single parent	-0.074	-0.069	-0.016
	(0.044)	(0.044)	(0.039)
Controls:			
Age & sex	No	Yes	Yes
GPA (elementary)	No	No	Yes
Observations	350	350	350
R-squared	0.117	0.121	0.303

Table A5: Gaps by SES categories at the time of tracking. Coefficients are OLS estimates using location FEs with White robust standard errors in parentheses. Sample: low and high SES control groups. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

	High track attendance at age 17 (SOEP, 10/11th grade)	
	(1)	(2)
High SES dummy	0.351***	
	(0.021)	
Low income dummy		-0.111***
		(0.022)
Low education dummy		-0.354***
		(0.022)
Single parent dummy		-0.013
		(0.023)
Observations	2,195	2,195
R-squared	0.101	0.165

Table A6: SES gaps at age 17 in the SOEP (grade 10/11). Coefficients are OLS estimates with White robust standard errors in parentheses. Track attendance is measured when participants are 17 years old, respectively. In the SOEP, track attendance is not observable at an earlier point in time. The sample includes the SOEP waves 2011-2018. All variables follow the definitions described in sections 2.2 and 2.3. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

	High track attendance in 5th grade		
	(1)	(2)	(3)
Mean low SES:		0.488	
Treatment dummy	0.110**	0.116**	0.110**
	(0.052)	(0.052)	(0.047)
Controls:			
Age & sex	No	Yes	Yes
GPA (elementary)	No	No	Yes
Observations	399	399	399
R-squared	0.110	0.118	0.302

Table A7: Causal effect of mentoring on high track attendance. Coefficients are OLS estimates using strata FEs with White robust standard errors in parentheses. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

	High track attendance in 9/10th grade		
	(1)	(2)	(3)
Mean low SES:		0.508	
Treatment dummy	0.104**	0.110**	0.103**
	(0.052)	(0.052)	(0.049)
Controls:			
Age & sex	No	Yes	Yes
GPA (elementary)	No	No	Yes
Observations	383	383	383
R-squared	0.143	0.153	0.277

Table A8: Causal effect of mentoring on high track attendance in grade 9/10. Coefficients are OLS estimates using strata FEs with White robust standard errors in parentheses. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

	GPA 4th grade	GPA 5th grade	GPA 9/10th grade
	(1)	(2)	(3)
Treatment dummy	-0.008	-0.154	-0.040
	(0.103)	(0.111)	(0.111)
Conditional on track	No	Yes	Yes
Observations	389	378	365
R-squared	0.101	0.104	0.077

Table A9: Effects on GPA. GPA is standardized by grade and higher values indicate better grades. Coefficients are OLS estimates with White robust standard errors in parentheses, using strata FEs. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

	"I enjoy going to school" (standardized)					
	Full sample		Only low track		Only high track	
	(1)	(2)	(3)	(4)	(5)	(6)
High track	0.297***	0.268**				
	(0.097)	(0.111)				
Overrule low recom.		-0.045				
		(0.162)				
Overrule high recom.		-0.146				
		(0.174)				
Treatment group			0.091		-0.007	
			(0.167)		(0.138)	
High SES				0.083		-0.015
				(0.394)		(0.137)
Observations	465	465	167	125	204	205
R-squared	0.022	0.024	0.003	0.009	0.010	0.003

Table A10: Enjoyment of schooling by groups and tracks. Coefficients are OLS estimates using location FEs with White robust standard errors in parentheses. Data collection at grade 8/9. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

	High track attendance					
	in 5th grade			in 9/10th grade		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment dummy	0.110**	0.106**	0.115*	0.104**	0.104**	0.107*
	(0.052)	(0.052)	(0.061)	(0.052)	(0.052)	(0.063)
IPW: attrition	No	Yes	No	No	Yes	No
IPW: representative	No	No	Yes	No	No	Yes
Observations	399	399	399	383	383	383
R-squared	0.110	0.112	0.131	0.143	0.145	0.166

Table A11: High track attendance in grade 5 and 9/10. Coefficients are OLS estimates using strata FEs with White robust standard errors in parentheses. Inverse probability weights (IPW) are used to check for biases due to (potential) selective attrition and (potential) non-representative SES distributions, details are provided in Appendices B.3.2 and B.3.1. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

	High track attendance			
	Grade 5 bounds		Grade 9/10 bounds	
	Lower	Upper	Lower	Upper
	(1)	(2)	(3)	(4)
Treatment Dummy	0.104	0.112	0.088	0.129
Number of obs.	590	590	590	590
Number of selected obs.	399	399	383	383

Table A12: Treatment effect bounds for high track attendance at grade 5 and grade 9/10. The bounds are estimated using the trimming procedure suggested by Lee (2009), using the implementation by Tauchmann (2014). Displayed bounds are weighted averages of location-specific bounds.

	High track attendance					
	in 5th grade			in 9/10th grade		
	(1)	(2)	(3)	(4)	(5)	(6)
LATE	0.152**	0.161**	0.152**	0.145**	0.153**	0.143**
	(0.072)	(0.072)	(0.065)	(0.072)	(0.072)	(0.067)
Controls:						
Age & sex	No	Yes	Yes	No	Yes	Yes
GPA (elementary)	No	No	Yes	No	No	Yes
Observations	399	399	399	383	383	383
R-squared	0.088	0.094	0.281	0.134	0.144	0.268

Table A13: Local average treatment effect (LATE) analysis using random group assignment as instrument for actual treatment. Coefficients are two-stage least squares (2SLS) estimates using random assignment as an instrument for actual treatment, using strata FEs. White robust standard errors in parentheses. In the sample in columns (1)-(3), 105 of the 145 children who we intended to treat were actually matched with a mentor. In the sample in column (4)-(6), 98 of the 137 children who we intended to treat were actually matched with a mentor. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

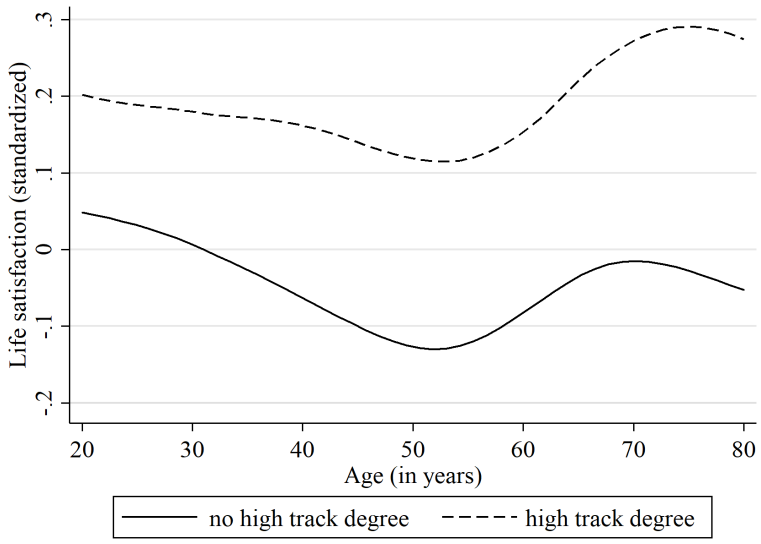

Figure A1: Life satisfaction by track degree over the life-cycle. Kernel-weighted local polynomial smoothing graph (bandwidth: 5 years) based on data of the SOEP (2019). Life satisfaction is measured on a 11-point Likert scale. Average gap amounts to 0.225 standard deviations ($p < 0.01$, using clustered SEs at the individual level, 525,712 observations of 67,828 individuals).

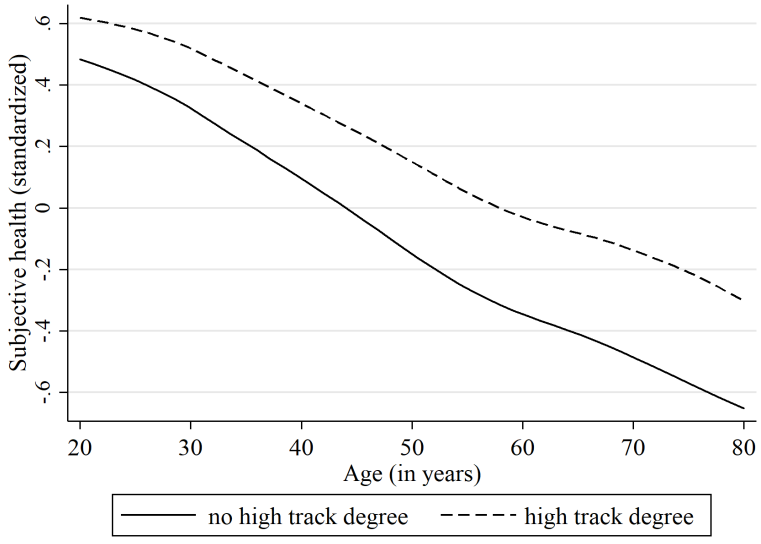

Figure A2: Subjective health by track degree over the life-cycle. Kernel-weighted local polynomial smoothing graph (bandwidth: 5 years) based on data of the SOEP (2019). Subjective health is measured on a 5-point Likert scale. Average gap amounts to 0.355 standard deviations ($p < 0.01$, using clustered SEs at the individual level, 444,748 observations of 63,001 individuals).

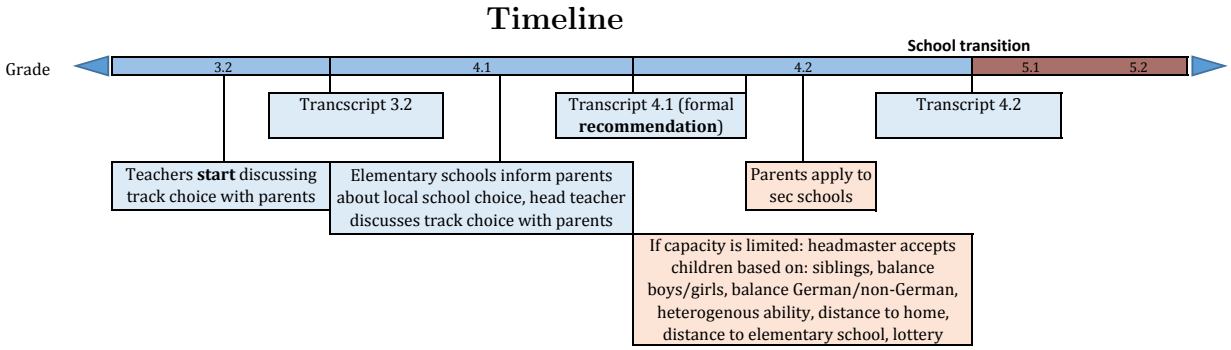

Figure A3: Timeline of the track choice and transition process.

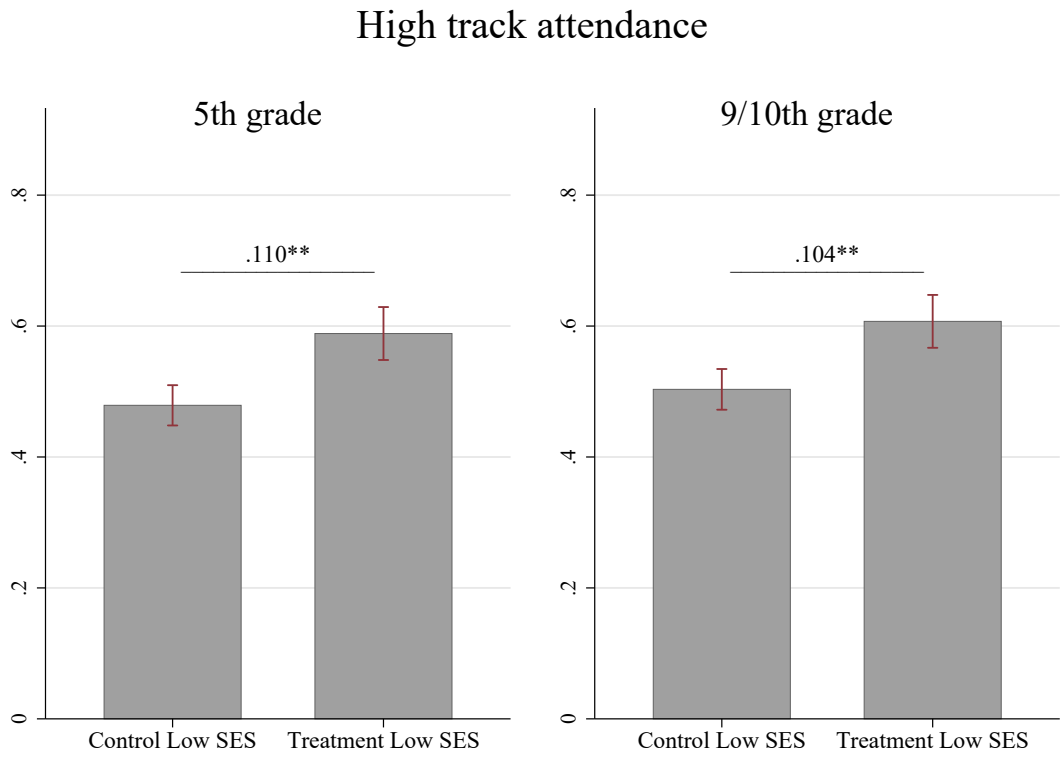

Figure A4: The graph displays the fraction of children who attend the highest track by treatment assignment category in grade 5 and 9/10. Error bars reflect standard errors. Results correspond to a linear probability model, for details see table A11. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

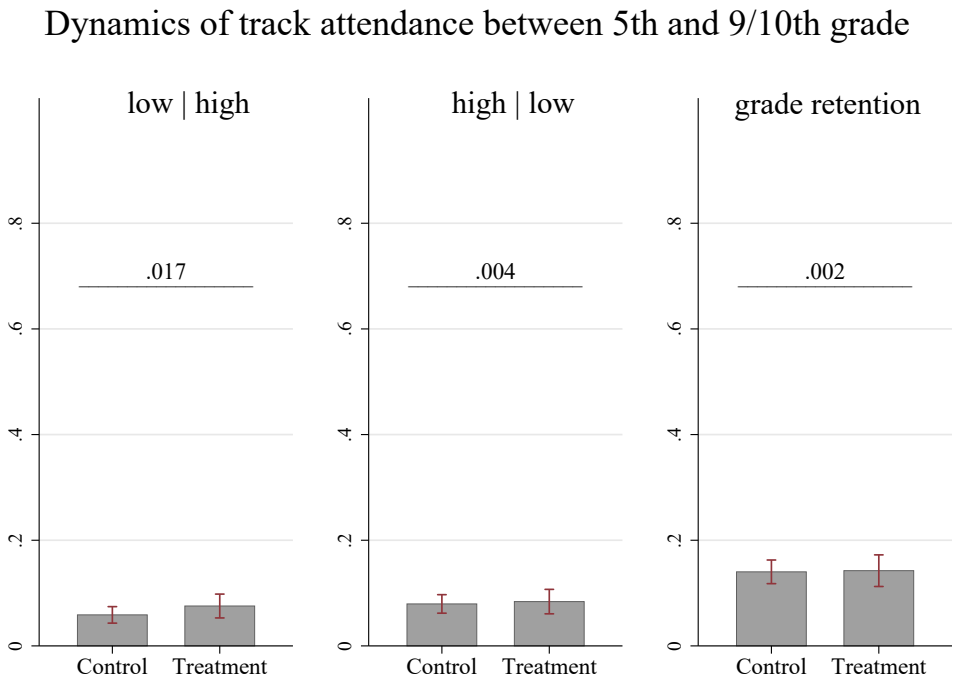

Figure A5: Long-run results: dynamics & grade retention. Error bars reflect standard errors. Results correspond to a linear probability models, using strata FEs. Displayed numbers indicate the differences between the two groups. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

B Additional Descriptions

B.1 Track definition

In Germany, all students jointly go to elementary school for at least four years. After elementary school, usually at age 10, students enter one of several secondary education programs. Education programs which lead to an upper secondary school degree, i.e., that qualify for university studies (German *Abitur*) are labeled "high track". Programs that award degrees which do not qualify for university studies are classified as "low track".[29] Children are categorized as high track students if they attend:

- Upper secondary school (Gymnasium, grades 5-12 or 13). This school type prepares students for university studies and awards them with an upper secondary school certificate (the Abitur) which allows them to enroll at university.
- "Advanced courses" in comprehensive schools (Gesamtschule). Comprehensive schools award several school leaving degrees. During the course of attending comprehensive schools students are, based on previous performance, successively sorted into basic and advanced courses. Attending advance courses in most subjects and having good grades allows students to attend an upper-secondary-program in grades 11 to 13 and gain an upper secondary school certificate (Abitur). The exact rules and details are determined by law, the so called "Ausbildungs- und Prüfungsordnung Sekundarstufe I" (APO-S I). In grade 5, we label comprehensive school students as being on the high track if their grades allow them to attend advanced courses in the following years. In grade 9/10 students are labeled as being on the high track if they attend at least three advanced courses and have good grades, i.e., allowing them to attend the upper-secondary-program in grades 11 to 13 leading up to the Abitur (see APO-S I).

B.2 Data collection

Recruiting Using official registry data we obtained more than 95% of the addresses of families living in the German cities of Bonn and Cologne who had children of ages seven to nine when the study started (October 2011). Offers to take

[29] These programs include lower and intermediate secondary schools (Hauptschule and Realschule), schools for children with special needs (Sonderschule) and comprehensive schools (basic course program). Seven students are not categorized as they attend schools with special pedagogical concepts without tracking.

part in the study were sent to all families with children born between 09/2003 and 08/2004 and to one third of the families with children born between 09/2002 and 08/2003. Dates were chosen such that children of the younger cohort were typically in second grade. In summer 2011, families were informed via postal mail about the study, including two waves of interviews and, in particular, the potential participation in a mentoring program. Parents were informed that participation in the mentoring program was not guaranteed due to limited capacity. If families were interested in participating and if their children were fluent in German, they were asked to sign a non-binding letter of intent to take part in the interviews and the mentoring program, and to send back a short questionnaire concerning socio-economic characteristics of the household.

Interviews Before and after the intervention, children and their mothers were interviewed by trained interviewers according to the same protocol. For pre- and post-treatment data collections, interviews took place in central locations in either Bonn or Cologne, Germany. The interviews and experiments were conducted by trained university students (mostly graduates) of psychology or education science. From 2014 (2-year follow-up) onward, yearly interviews of the same families took place at their homes. Interviews and experiments were conducted by experienced and specially trained interviewers of TNS Infratest within the official framework of the innovation sample of the German Socio-Economic Panel (SOEP-IS) (Wagner et al., 2007). At no point in time interviewers were informed about the purpose of the study or the group assignment of the participating families (treatment/control or high/low SES). The interviews and experiments were conducted according to a detailed protocol. For participation in the interview, parents received 35 Euro in the pre-treatment wave and 45 Euro in the following waves.

B.3 Robustness checks: weighting

B.3.1 Weighting scheme 1: attrition

Our core samples comprises all children for whom schooling information and especially track attendance in grade 5 (and in grade 9/10) are available. These comprise 70% (67%) of the baseline sample. In order to check for a potential bias by selective attrition related to treatment status and child school performance, we re-weight the observed data using inverse probability weighting (IPW). Weights are estimated from a Probit model of a binary selection indicator (indicating whether track information is available) regressed on treatment assignment and baseline GPA and

their interaction. The re-weighted main results are displayed in table A11. As the effects sizes are very similar to the unweighted effect sizes, we conclude that selective attrition does not bias our results.

B.3.2 Weighting scheme 2: representative population

In order to assess the representativeness of our results, we compare our sample to a representative sample of families in Germany. The comparison comprises of all three dimensions of parental SES: income, education and single parent status, as well as children's elementary school GPA. As a reference, we make use of data from the German Socio-Economic Panel (SOEP) (Wagner et al., 2007). The SOEP is comprised of yearly data on more than 20,000 individuals in more than 10,000 households. We compare the children in our sample with children in the same age-range as in the SOEP (2019). In order to compare the samples along the SES dimensions that we used in the sampling process, we elicited SES in the same way as in the SOEP (for details, see section 2.2).
In order to check if deviations from representativeness, in terms of the SES dimensions and child ability, affect our results, we re-weight our data based on propensity scores that model the probability of being in our sample. For this purpose, we estimate propensity scores based on a Probit model that includes dummies of all SES dimensions, elementary school GPA and their interactions. The re-weighted main results are displayed in table A11. The resulting effect sizes are very similar to those from the unweighted regression, suggesting that selective participation does not bias our results.

C A Decomposition of the Mentoring Effect

A child is observed in a high track program in grade 5 ($h = 1$) whenever one of the following conditions holds:

(i) the child receives a high recommendation ($r = 1$) and and her parents adhere to it ($h = 1|r = 1$) or

(ii) the child receives a low recommendation ($r = 0$) but her parents decide to overrule ($h = 1|r = 0$) it.

In the absence of the treatment, we can thus write the probability of being in the high track in grade 5 as:

$$\begin{aligned} P(h=1) &= P(h=1, r=1) + P(h=1, r=0) \\ &= P(h=1|r=1) * P(r=1) + P(h=1|r=0) * P(r=0) \end{aligned}$$

Restricting the sample to low SES children, each of the above objects can be calculated from our data for the treatment and control groups ($t = \{0, 1\}$), respectively. Then, taking finite differences with respect to the treatment yields:

$$\begin{aligned} \Delta_\tau P(h=1) = \; & \Delta_\tau P(h=1|r=1) * P(r=1|t=0) + P(h=1|r=1, t=0) * \Delta_\tau P(r=1) \\ & + \Delta_\tau P(h=1|r=1) * \Delta_\tau P(r=1) \\ & + \Delta_\tau P(h=1|r=0) * P(r=0|t=0) + P(h=1|r=0, t=0) * \Delta_\tau P(r=0) \\ & + \Delta_\tau P(h=1|r=0) * \Delta_\tau P(r=0), \end{aligned}$$

where Δ_τ denotes the effect of treatment assignment, e.g., $\Delta_\tau P(h=1) = P(h=1|t=1) - P(h=1|t=0)$. After collecting terms, and since $\Delta_\tau P(r=0) = -\Delta_\tau P(r=1)$, we can rewrite the overall intention to treat effect as being approximately equal to (i.e., disregarding the interaction effects):

$$\begin{aligned} \Delta_\tau P(h=1) \approx \; & \overbrace{\Delta_\tau P(r=1) * [P(h=1|r=1, t=0) - P(h=1|r=0, t=0)]}^{\text{child recommendation effect}} \\ + \; & \underbrace{\Delta_\tau P(h=1|r=1) * P(r=1|t=0) + \Delta_\tau P(h=1|r=0) * P(r=0|t=0)}_{\text{parental decision effect}}, \end{aligned}$$

In line with the above, the overall effect can then be decomposed into two parts. A part that reflects a change in children's recommendations (child recommendation effect) and a part that reflects a change in parental decision-making conditional on track recommendation (parental decision effect).

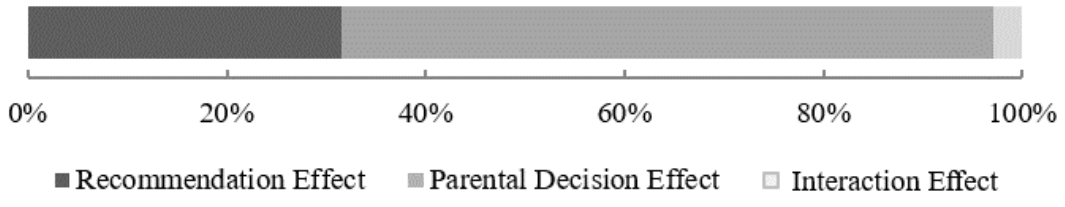

Figure C1: Estimates of the decomposition as described in the above equation. The bar represents the total treatment effect.

	High track in grade 5	High Recom.	High in grade 5 given	
			low recom.	high recom.
	(1)	(2)	(3)	(4)
Mean Low SES Control:	0.488	0.547	0.139	0.777
Treatment dummy	0.106**	0.057	0.038	0.096*
	(0.053)	(0.052)	(0.060)	(0.053)
Observations	399	399	173	226
R-squared	0.013	0.003	0.015	0.018

Table C1: Decomposition. Coefficients are OLS estimates using location FEs with White robust standard errors in parentheses. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

We compute each of the above probabilities in our data and collect terms to obtain the part of the program effect that can be ascribed to children (recommendation effect) and parents (parental decision effect), respectively. Figure C1 displays the results of this decomposition (estimates are shown in table C1). The dark gray bar indicates that roughly one third of the effect on high track attendance arises due to the positive impact of the mentor on the children and the resulting change in teacher recommendations. Moreover, roughly two-thirds of the explained program effect can be ascribed to the change in parental decision-making.

D SES Differences and Opportunity Gaps at the two Decision-Making Stages

	Teacher recommendation			**Parental decision making**					
	Recommendation for high track			High track given low recom.			High track given high recom.		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mean low SES:	0.547			0.139			0.777		
High SES dummy	0.286***	0.288***	0.186***	0.190	0.201*	0.197*	0.111**	0.112**	0.103*
	(0.051)	(0.051)	(0.044)	(0.121)	(0.117)	(0.114)	(0.054)	(0.053)	(0.052)
GPA (std.)			0.258***			0.073***			0.138***
			(0.019)			(0.028)			(0.037)
Additional controls:									
Age & sex	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Observations	350	350	350	131	131	131	219	219	219
R-squared	0.070	0.079	0.357	0.036	0.081	0.118	0.019	0.036	0.101

Table D1: The role of children's recommendations and parental decisions. Coefficients are OLS estimates using location FEs with White robust standard errors in parentheses. GPA is the standardized elementary school GPA, coded such that higher values indicate better performance. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

To unveil the extent to which the decision-making process translates into inequalities of opportunity, this section displays socio-economic differences in children's recommendations and parental decision-making, respectively. The results in Table D1 show SES differences at both decision-making stages. First, Control Low SES children are overall 28.6 percentage points less likely to receive a high track recommendation than Control High SES children. Moreover, this difference reduces by 35% to 18.6 percentage points after we control for child elementary school performance. These findings again point towards discrimination against low SES children as their parental background seems to weigh more heavily than school performance for a high teacher recommendation. Besides, equality of opportunity differences prevail also at the parental decision-making stage. As indicated in columns 4-9 of table D1, low SES parents are about 20 percentage points less likely to overrule a low recommendation (see columns 4-6) upwards. Moreover, they are around 10 percentage points less likely to follow along with a high recommendation (see columns 7-9), indicating that some low SES parents choose the low track, although their children have received high track recommendation. Throughout, child performance at school seems largely unimportant for parental decision-making. This suggests that any differences between the unconditional SES gap and the conditional SES gap observed

in figure 1 are driven by performance-induced changes in the teacher recommendation while parental decision-making seems largely insensitive to differences in child school performance.