# Can Mentoring Alleviate Family Disadvantage in Adolescence? A Field Experiment to Improve Labor-Market Prospects[*]

Sven Resnjanskij, Jens Ruhose, Simon Wiederhold, Ludger Woessmann, and Katharina Wedel[†]

## Abstract

We study a mentoring program that aims to improve the labor-market prospects of school-attending adolescents from disadvantaged families by offering them a university-student mentor. Our RCT investigates program effectiveness on three outcome dimensions that are highly predictive of later labor-market success: math grades, patience/social skills, and labor-market orientation. For low-SES adolescents, the mentoring increases a combined index of the outcomes by over half a standard deviation after one year, with significant increases in each dimension. Part of the treatment effect is mediated by establishing mentors as attachment figures who provide guidance for the future. Effects on grades and labor-market orientation, but not on patience/social skills, persist three years after program start. By that time, the mentoring also improves early realizations of school-to-work transitions for low-SES adolescents. The mentoring is not effective for higher-SES adolescents. The results show that substituting lacking family support by other adults can help disadvantaged children at adolescent age.

Keywords: mentoring, disadvantaged youths, adolescence, school performance, patience, social skills, labor-market orientation, field experiment

JEL code: I24, J24, H52

April 12, 2023

[*] We gratefully acknowledge comments and suggestions from the editor, three anonymous referees, Flavio Cunha, Chris Flinn, Fabian Kosse, Philip Oreopoulos, and seminar participants at Bamberg, Chicago, Essen, Essex, Heidelberg, Ispra, Lund, Munich, Torino, Stavanger, ifo, EEA, EALE, IIPF, IWAEE, AEDE, German Economic Association, and CESifo Economics of Education Area. We are grateful to Franziska Hampf for exceptional collaboration in the early survey execution and data collection and to Hanna Brosch, Vera Freundl, Sarah Kersten, Franziska Kugler, Katja Skubik, and Anna Wurm for excellent data-collection and research assistance. Financial support by the Wübben Foundation, Jacobs Foundation, Porticus (through Founder Association for German Science), and the Commissioner of the Federal Government for Migration, Refugees, and Integration is gratefully acknowledged. The contribution by Woessmann is part of German Science Foundation project CRC TRR 190 and Leibniz Competition project SAW 2019. The project was granted IRB approval by the Ethics Commission of the Department of Economics of the University of Munich (Project 2016-02). A pre-analysis plan that specified the two-cohort sampling design, the three outcome dimensions, and disadvantaged adolescents as the target group of the evaluation was contained in the grant application registered with the funding foundations on May 12, 2015. The RCT is registered at the AEA RCT Registry (AEARCTR-0006613).

[†] Resnjanskij: ifo Institute at the University of Munich and CESifo, resnjanskij@ifo.de; Ruhose: Kiel University, CESifo, and IZA, ruhose@economics.uni-kiel.de; Wiederhold: Halle Institute for Economic Research (IWH), University of Halle, ifo Institute, CESifo, and IZA, simon.wiederhold@iwh-halle.de; Woessmann: University of Munich, ifo Institute; Hoover Institution, Stanford University; CESifo, and IZA, woessmann@ifo.de; Wedel: ifo Institute at the University of Munich, wedel@ifo.de.

## 1. Introduction

The persistence of inequality across generations is a major concern worldwide (e.g., Black and Devereux (2011); Corak (2013); Autor (2014); Alvaredo et al. (2018)), also in countries that maintain an extensive social welfare system.[1] A defining characteristic of children from disadvantaged backgrounds is that they lack the powerful family support that other children receive by the "accident of birth" (Heckman (2008), p. 289). Therefore, policies aimed at helping disadvantaged children face dire limitations as neither schools nor family-targeted programs can fully substitute or change parents. Existing evidence suggests that interventions stand a good chance to succeed if they aim to compensate for lacking family support already early in life (e.g., Cunha et al. (2006); Almond, Currie, and Duque (2018); García et al. (2020); Kosse et al. (2020)). By contrast, later interventions in schools or labor markets have proven much less successful in helping disadvantaged youths (e.g., Cunha et al. (2006)). However, little attention has been given to later interventions that provide personal support from other adults. This is the approach followed by numerous mentoring programs that aim to help adolescents from disadvantaged backgrounds by assigning them a mentor who can provide them with support that their family environment is not able to provide.

In this paper, we report results of a field experiment that evaluates whether mentoring can help disadvantaged adolescents to improve their school performance and skill development to achieve long-term success on the labor market. We study a nationwide German mentoring program that offers adolescents from disadvantaged families in low-track secondary schools a voluntary university-student mentor to prepare them for a successful transition into professional life. The core of the program consists of regular mentor-mentee meetings focused on developing the adolescents' individual potential, career orientation, school assistance, and leisure activities. The program is set up for one year, with a possible extension of mentoring meetings up to a maximum of two years. It is organized as a social franchise with a centralized concept and support structure that is implemented in 42 self-governing locations.

To evaluate the impact of the program, we conducted a randomized controlled trial (RCT) among 308 adolescents in 10 city locations serving 19 schools in two cohorts. At program start, the adolescents are on average 14 years old. Randomization relied on local program

[1] For example, in Germany it takes six generations for those born in low-income families to approach the mean income in their society, longer than in the United States (five) and the OECD average (4.5) (OECD (2018)).

oversubscription. After surveying all adolescents before program start, we implemented a pair-wise matching design with rerandomization that ensures balancing of baseline observables across treatment and control groups. We invested substantial effort to reach participants one year after program start, including more than 100 person-trips to participating schools for data collection in a school context. As a result, we achieve a recontact rate of 98.7 percent (304 of the 308 participants) one year after program start and – tracking participants further – 88.3 percent three years after program start.

Our main analysis investigates program effectiveness on three outcome dimensions that are highly predictive of adolescents' long-term labor-market success:[2] math grades as a mostly cognitive component,[3] patience and social skills as a behavioral component, and labor-market orientation as a volitional component. We combine the three components into one index of labor-market prospects to capture the overall program effect and to alleviate concerns of multiple hypothesis testing. Throughout, our analysis separates between adolescents from highly disadvantaged backgrounds (low socioeconomic status (low-SES)) who are the main target group of the program and higher-SES adolescents who are also eligible to participate. Our baseline model splits the sample roughly half and half based on a multidimensional SES measure that combines information on books at home, parental education, single-parent status, and first-generation migrant status. Results are consistent across the four sub-components.

We find that the highly disadvantaged youths benefit strongly from participation in the mentoring program. At the end of the main program period one year after program start, program participation increases the index of labor-market prospects of low-SES adolescents by more than half a standard deviation, closing the initial gap in labor-market prospects to the higher-SES adolescents in the sample. In the preferred model with controls for the pre-treatment value of the outcome measure and a full set of randomization-pair fixed effects from the pair-wise matching, the intention-to-treat effect is 0.64 standard deviations. By contrast, the program does not significantly affect higher-SES adolescents, whose labor-market prospects are if anything lower due to program participation. The difference in the treatment effect between low-SES and

[2] In section 4.2 and Appendix E, we provide evidence for the labor-market relevance of each component.

[3] It is well established that school grades reflect both cognitive and non-cognitive skills (e.g., Borghans et al. (2016)). In the exposition, we emphasize the cognitive component in grades because the second sub-index of labor-market prospects, patience and social skills, directly incorporates non-cognitive skills. We regard the non-cognitive component in grades, which reflects pupils' personalities as assessed by teachers, as complementary to the patience and social skills index, which is based on adolescents' self-reports.

higher-SES adolescents is highly significant. Average program effects are significantly positive, but relatively modest in size.

Also for each of the three (cognitive, behavioral, and volitional) components, the mentoring program has a significant positive treatment effect for low-SES adolescents after one year, but an insignificant negative effect for higher-SES adolescents. For low-SES adolescents, school grades in math increase by 0.31 standard deviations. The program increases their index of patience and social skills by 0.47 standard deviations, mostly driven by the patience sub-component. The index of labor-market orientation rises by 0.44 standard deviations for low-SES adolescents. Overall, the mentoring program positively affects a range of outcomes that are important for long-term labor-market success, but have generally been thought of as difficult to change at adolescent age. Our results suggest that substituting lacking family support by other adults can help disadvantaged children not only in early childhood, but also in adolescence.

Mediation analysis suggests that successfully establishing an additional attachment figure with whom low-SES adolescents can talk about their future acts as a mediator of the treatment effect. Additional aspects of the mentor-mentee relationship that may facilitate the transition into professional life are that treated low-SES adolescents are more likely to perceive their mentors as an important source of information for occupational choice and to perceive schools as useful for future jobs. Together, these three mediators account for 37 percent of the overall treatment effect for low-SES adolescents. Descriptive analysis of information on the mentoring relationships in the treatment group indicates that there are no relevant SES differences in the frequency, duration, or content of the mentoring meetings that could account for the differences in treatment effects by SES. Thus, the mechanism analysis suggests that the mentoring is successful by establishing mentors as attachment figures who provide guidance for the future.

Results on the persistence of treatment effects beyond the program period, two and three years after program start, differ across the three outcome components. The treatment effect on math grades remains substantial until the end of low-track secondary school, although there is some fade out. The effect on labor-market orientation persists well beyond the duration of the mentoring program and even increases over time. By contrast, treatment effects on patience and social skills are no longer detectable two and three years after program start. However, given the other results, it seems plausible that the mentoring has shifted adolescents' patience in a period critical for setting the course for their future careers.

Looking at early realizations of labor-market outcomes, program participation increases the likelihood that low-SES adolescents pursue an apprenticeship three years after program start by 29 percentage points. Completing an apprenticeship would imply a highly successful transition into professional life for most program participants in the German context.[4] However, the possibility to investigate treatment effects on actual labor-market outcomes is limited by the fact that a large share of adolescents is still in school even three years after program start. But the large positive treatment effect on the probability to pursue an apprenticeship is strong evidence of program effects on adolescents' emerging labor-market outcomes. Treated low-SES adolescents are also more satisfied with their current situation than those in the control group three years after program start. Overall, the mentoring program seems to have set the stage for disadvantaged youths for a successful transition into working life.

Our paper contributes to the literature on mentoring interventions to help disadvantaged youths. Despite the broad prevalence of mentoring programs for adolescents, there is surprisingly little evidence on their causal effect on labor-market prospects. Recent experimentally studied interventions tend to combine mentoring with other elements such as financial incentives, academic tutoring, and additional educational services into comprehensive support programs, making it hard to assign treatment effects to any specific component. For example, the Quantum Opportunity Program studied by Rodríguez-Planas (2012) combines mentoring with additional educational services and financial incentives. In the programs studied by Heller et al. (2017), mentoring is just one component in a curriculum of many activities focused on cognitive-behavioral therapy in group sessions. The Pathways to Education program studied by Oreopoulos, Brown, and Lavecchia (2017) and Lavecchia, Oreopoulos, and Brown (2020) is a comprehensive support program that integrates mentoring, daily tutoring, group activities, and various financial incentives. The intervention we study is a pure mentoring program that allows us to assess the effectiveness of a relatively low-intensity, low-cost support program.

Most of the available studies on pure mentoring programs are non-experimental (see DuBois et al. (2002), Rhodes (2008), Eby et al. (2008), and Rodríguez-Planas (2014) for overviews indicating modest average program effects). The main exception is the Big Brothers Big Sisters

[4] Compared to no professional qualification, completing an apprenticeship is associated with substantially lower unemployment (4.2 vs. 19.1 percent, IAB (2017)) and 31 percent higher lifetime earnings (Piopiunik, Kugler, and Woessmann (2017)) on the German labor market.

Program evaluated for 9- to 16-year-old children, which has been found to reduce drug abuse and school absenteeism and improve family relationships in an outside-school delivery with adult mentors (Grossman and Tierney (1998)) and to improve academic performance, but not effort, self-worth, family relationships, or problem behavior in a within-school delivery with mostly high-school student mentors (Herrera et al. (2011)). However, the program had no particular aim to improve labor-market prospects, an outcome of core interest in the economics literature that is the goal of our studied mentoring program and the subject of our evaluation.[5]

The remainder of the paper is structured as follows. The next two sections describe the mentoring program and the implementation of our RCT. Section 4 describes the main variables and section 5 discusses the empirical strategy. Sections 6 and 7 present our main results on program effects on labor-market prospects at the end of the main program period and an analysis of mechanisms. Section 8 reports results on the persistence of program effects up to three years after program start and on early realizations of labor-market transitions. Section 9 provides a cost-benefit analysis indicating that the program is highly cost-effective and a scalability analysis showing that the program's reach could be substantially expanded. Section 10 concludes.

## 2. The Mentoring Program

We study the effectiveness of one of the largest one-to-one mentoring programs for disadvantaged youths in Germany. The program, called *Rock Your Life!*, was founded by a group of university students in 2008. It is offered in 42 cities across Germany (and ten cities in Switzerland and the Netherlands) and has established more than 7,000 mentoring relationships since its foundation (Rock Your Life! (2020)). The mentees are adolescents from lowest-track secondary schools (*Hauptschule* or equivalent in the German system where different types of schools cater for different academic levels) who are assigned a university student as a mentor. The main goal of the program is to prepare the adolescents for a successful transition into professional life. The program aims at providing career guidance, establishing career visions, and fostering self-esteem and trust in the mentees' own skills and abilities. Each mentoring pair is free to choose the content of their relationship, striving for at least bi-weekly meetings. While the mentoring activities include joint spare-time activities such as going to the cinema or the zoo,

[5] Two recent mentoring studies in elementary-school contexts investigate effects on prosociality (Kosse et al. (2020)) and truancy (Guryan et al. (2021)).

mentors may also counsel mentees how to cope with stressful situations at school or in the family, provide occupational orientation, and assist in the job application process.

The program is organized as a franchise system of self-governing university societies in each participating university town, which are responsible for operating and organizing the mentoring program. The societies recruit university students to act as mentors on a voluntary basis. They use screening devices to select suitable candidates from the pool of applying university students, typically based on certificates of good conduct and personal interviews. Because the mentoring relationships are meant to last for at least one year with a possible extension of up to two years, it is common that each admitted student serves as a mentor only once during the society membership. An umbrella organization, organized as a non-profit holding, coordinates and oversees the activities of the mentoring sites, represents the mentoring program to the outside, and is responsible for strategic decisions on the future direction of the overall program. The holding provides standardized training courses for the mentors, counseling of mentors on how to run the mentoring relationship, and training on how to organize the university societies. The program relies on funding from foundations and other social investors.

The program is targeted at pupils in eighth and ninth grade. It is meant to run through the final two years before leaving the lowest-track secondary schools.[6] In each participating city, the university society typically selects two to four low-track schools in disadvantaged neighborhoods to recruit adolescents for program participation. Compared to the average adolescent in Germany, targeted adolescents are disadvantaged because they usually visit a secondary school of the lowest academic track and often have a migrant background. However, there is no screening of potential participants within the participating low-track schools.

The initiation of the mentoring relationship follows a predetermined structure. In the first step, university-student officials of the society visit participating schools located in their city to introduce the program in front of an entire grade level. All adolescents in the grade can apply for the program. While teachers and principals are also free to recommend adolescents who they feel would benefit most from the program, program admission is not based on teacher recommendations. Interested adolescents receive information material for themselves and their parents, as well as consent forms to be signed by parents with which they apply to the program. During a *Kick-Off* training, participating adolescents then get to know the mentors in a round of

[6] Low-track schools in most German states used to last until grade nine but mostly extend to grade ten by now.

introduction and the one-to-one mentoring relationships are formed.[7] The default is that adolescents are matched to mentors based on mutual preferences directly after the introduction phase; eventually, each mentee gets assigned a mentor.[8] Matches of female mentees to male mentors are not allowed. While some sites allow matches of male mentees to female mentors, most allow only same-sex matches.

## 3. The RCT

To evaluate whether the mentoring program is effective in improving adolescents' labor-market prospects, we designed and implemented a field experiment. This section describes the setup of the RCT (section 3.1), the baseline survey and randomization before program start (section 3.2), and the follow-up surveys conducted one to three years after program start (sections 3.3 and 3.4).

### 3.1 Setup

In designing the RCT, we aimed to exploit the fact that oversubscriptions frequently occurred in the nationwide expansion phase of the mentoring program where sites generally aimed to increase the number of participants and new sites were regularly founded. We randomly assigned program applicants to a treatment group offered to participate in the mentoring program and a control group. Adolescents in the control group did not have the opportunity to participate in the mentoring program but were offered an incentive not related to the content of the mentoring program to mitigate discouragement effects.[9]

Our pre-analysis plan specified a two-cohort sampling design. Sites were selected for participation in the RCT based on criteria designed to represent the target population of the mentoring program and to avoid cream skimming by the program (e.g., Heckman (2020)). All contacted sites and schools agreed to participate.

---

[7] The program includes three compulsory trainings, each consisting of one joint day for mentors and mentees and one day just for mentors. The *Kick-Off* training lays the foundation for the relationships. The *Job-Coach* training takes place after three to six months and focuses on career orientation and potential development. In the *Your-Way* training after one year, mentors and mentees reflect on what has already been achieved during the relationship.

[8] Some sites use a different allocation mechanism, e.g., assigning a higher weight to the mentees' than the mentors' choice. In rare cases, mentors are allocated to mentees by officials from the mentoring site.

[9] These incentives were mainly one of the following: cinema ticket, Christmas party, one-day job training, or firm visit. To the extent that participation in these activities had an effect on labor-market orientation, this would introduce downward bias in our estimates. In practice, however, demand for the incentives was typically very low.

In total, 11 mentoring sites in 12 cities spread across Germany participated in the evaluation. The main data collection for the baseline survey took place between October 2016 and May 2017 in the different sites of the first cohort and one year later in the second cohort (see Figure 1 and Appendix B).[10] Randomization was performed directly after the baseline survey in each site, and the program started shortly afterwards. About one, two, and three years after program start (for each site and cohort), we fielded follow-up surveys to collect outcome data. In addition, we collected administrative school data until the end of low-track secondary school.

To circumvent randomization bias (e.g., Heckman (2020)), our RCT did not alter any elements of the program or the preselection of adolescents who opted into the program. We were neither involved in nor did we influence which schools were targeted by the mentoring sites, how principals, teachers, and pupils were addressed, and how university students acting as potential mentors were selected and admitted. Moreover, mentors were not systematically informed by the mentoring sites that the program is subject to an evaluation. Of course, in sites with program oversubscription, our study design enforced a randomized allocation into treatment.[11]

### 3.2 Baseline Survey and Pair-Wise Randomization

Before program start, we collected baseline data for all applying adolescents in which we surveyed basic demographic, socioeconomic, and family characteristics, as well as measures of school performance, behavior, and economic preferences. Baseline data were collected in participating schools through a pen-and-paper survey administered by members of the project team.[12] Overall, 442 adolescents completed the baseline survey.

We use a pair-wise matching design with rerandomization to assign applicants into treatment and control groups within pairs of statistical twins. Randomization was implemented separately for each site, so that local environments are perfectly balanced. The matching was performed to minimize within-pair distances in a vector of matching variables (gender, classroom, and math and German grades) observed in the baseline survey. Performing 1,000 within-pair randomization replications, we chose the iteration that provided the best balancing

[10] The first cohort also includes two pilot studies fielded in November 2015 and June 2016.

[11] In the years before the RCT, oversubscription was also common and usually handled on a case-by-case basis, such as first-come-first-serve, recommendations by teachers or local program administrations, or coin flip.

[12] Questionnaires were filled by respondents in the classroom or another room (e.g., assembly hall) offered by the school. Interviewers made sure that sufficient space and/or visual protection existed between respondents to prevent any interaction between them while filling the questionnaires. The baseline questionnaire had been tested extensively prior to the evaluation in a school in Munich to ensure that pupils properly understood the questions.

for a set of eleven baseline variables (see Appendix C for details). The pair-wise matching approach has three desirable features compared to simple or stratified randomization (e.g., Bruhn and McKenzie (2009); Morgan and Rubin (2012); Imbens and Rubin (2015)). First, it provides better balancing properties within small samples. Second, treatment effects can be more efficiently estimated due to the inclusion of pair fixed effects. Third, it is possible to preserve internal validity of estimates in case of sample attrition if a participant leaves the sample by also dropping the statistical twin. The outcome of the randomization was reported to the mentoring site before mentoring relationships were initially formed.[13]

We could randomize applicants into treatment and control groups only if there was oversubscription of applicants (more applying pupils than available mentors) at the local level. However, not all participating sites achieved oversubscription because the number of applicants at each site is, to some extent, subject to natural variation.[14] At sites without oversubscription, randomization of program assignment was not feasible, and all applicants were treated.[15] As a consequence, our final estimation sample consists of 308 adolescents attending 19 different schools in 10 cities who were randomly assigned in matched pairs, 153 to the treatment group and 155 to the control group.[16]

### 3.3 First Follow-Up Survey One Year after Program Start

Our main analysis to evaluate the effects of the mentoring program on labor-market prospects focuses on the situation about one year after program start when the program ends for most adolescents and most of them are still exposed to the same school environment. The first follow-up survey was conducted similarly to the baseline, i.e., respondents filled the surveys in

[13] To avoid potential discouragement effects, the result of the randomization was not disclosed in front of classmates, but the holding sent decision letters to the applicants' home addresses by mail.

[14] We found no evidence for an effect of the evaluation on application decisions of adolescents. Participation in the evaluation was no prerequisite to apply for the program, and we communicated that the odds to obtain a slot in the program were independent of participation in the evaluation. In very few cases, applicants had to be included in the program before the random assignment because officials from the mentoring site or teachers felt that the respective applicant was in major need of the program. In these cases, we randomized the remaining individuals.

[15] Adolescents who could not be included in the randomization are similar to those in the randomized sample (Appendix B). Results are very similar when we add the adolescents in the non-randomized treatment group to the analysis (not shown), suggesting that the program effect does not systematically differ between sites with and without oversubscription.

[16] The number of observations in treatment and control groups can differ in cases of uneven numbers of applicants at a site. With uneven numbers, the final group in the pair-wise matching contains three applicants, one or two of whom are assigned to treatment (depending on whether one or two mentors remain).

their schools to maximize participation.[17] In the few cases where pupils were not present at school at the day of the survey, we either asked the teacher to hand out the survey questionnaire once the pupil returned to school or – if the pupil had moved to a different school – tried to contact the pupil ourselves by phone. In total, 94.5 percent of respondents whom we reached with the follow-up survey conducted the survey at school at the day of the interview, 1.7 percent conducted the survey at school at a different day, and 3.8 percent could be reached via phone.[18]

In addition to the survey information, we collected administrative information on pupils' school grades at baseline and in the follow-ups. These administrative data are available from pupils' report cards that are stored in the respective schools' archives.

One year after program start, we were able to achieve very high participation in the follow-up survey and coverage of the administrative data. For 304 out of the 308 adolescents in the randomized sample (98.7 percent), we have follow-up information either from the survey or from the report cards. Considered separately, the participation rate is 94.5 percent in the follow-up survey and 95.5 percent for the administrative follow-up information. This exceptionally high recontact rate is a result of the fact that we exerted substantial effort to organize the surveys in a school context, which entailed a total of more than 100 person-trips by our team members to schools to talk to principals and teachers, administer the surveys, and collect administrative data.

### 3.4 Additional Follow-Up Surveys Two and Three Years after Program Start

We also gathered information about the school and labor-market situation of the adolescents in two additional follow-up surveys two and three years after program start. The aim was to test whether first-year effects are persistent beyond the program period and to give a very early preview of labor-market transitions. A limitation is that pupils tend to attend different schools by this time, potentially compromising comparability of school outcomes. Comparison of outcomes is further impaired by the fact that some adolescents have entered the labor market while others still attend different (compensatory or continuing) types of schools.

---

[17] Treatment and control respondents were surveyed together. They were not aware that there were slightly different questionnaires for the two groups, as this was not announced and all questionnaires had the same cover page. All clarification questions by the respondents were answered individually by the interviewers, to make sure that any question regarding the mentoring program would not get noticed by respondents in the control group.

[18] Results are robust to adding survey mode fixed effects and to restricting the sample to participants who conducted the survey at school. Questionnaires completed at school at a different day were sent back to us by a contact person in the school (usually not the participants' main teacher) by regular mail.

The second follow-up survey took place about two years after program start in each cohort and site (Figure 1). A total of 261 (out of 308) adolescents participated in this survey, resulting in a recontact rate of 84.7 percent. We adjusted our sampling procedure to the fact that only about two thirds of adolescents still attend their original school in the second follow-up survey. Eventually, 50 percent of our completed surveys were collected in school, 38 percent via an online questionnaire, 11 percent by regular mail, and 2 percent by phone.

We conducted the third follow-up survey about three years after program start. In total, 272 adolescents participated in the survey (88.3 percent recontact rate. Due to the Covid-19 pandemic, the survey was carried out as an online questionnaire; 3 percent of respondents who could not be reached online were surveyed by phone. The information provided in the third follow-up survey illustrates the various pathways on which disadvantaged adolescents enter professional life. However, the survey also shows that the majority of adolescents (56 percent) is still in some type of school three years after program start.

We also collected administrative school data to the end of low-track secondary school. We obtained data on grades from report cards, available for 292 adolescents (94.8 percent), and on graduation outcomes, available for 300 adolescents (97.4 percent).

Attrition analyses in Appendix D indicate that attrition is very similar in treatment and control groups in the different follow-up surveys and is not selective with respect to baseline observables in the total sample, the low-SES subsample, or the higher-SES subsample.

## 4. Data and Variable Definitions

This section describes the measurement of adolescents' socioeconomic background as a potential source of treatment-effect heterogeneity (section 4.1) and the construction of our main outcome measure of labor-market prospects one year after program start (section 4.2).

### 4.1 Characterizing Socioeconomic Background

The mentoring program mainly targets highly disadvantaged adolescents. However, when analyzing the baseline survey data, we learned that a non-negligible share of participants has a family background that cannot necessarily be considered as *highly* disadvantaged. The mentoring program is active in low-track schools in disadvantaged neighborhoods in relatively large cities, each of which leads to a disproportionately high share of disadvantaged youths. However, the program does not implement any screening or selection of applying adolescents within the participating schools, leading to rather diverse family backgrounds among participants.

When planning the design of the RCT, our hypothesis was that mentoring is mainly successful for highly disadvantaged adolescents who are severely lacking family support. While the program might also be useful for labor-market prospects of more advantaged individuals, it may not have an effect if mentors do not contribute more than the adolescents' families already do. In fact, the effect may even turn negative if the mentoring crowds out more useful inputs offered by more advantaged families such as parental attachment or participation in other useful activities. Therefore, we investigate heterogeneous treatment effects by socioeconomic status (SES) throughout.

To identify the lack of family support, we classify adolescents as "low-SES" if at least one of the following three conditions holds: (i) lack of *educational* support: no university-educated parent and having few books at home;[19] (ii) lack of *economic or time* support: living with a single parent and having few books at home; (iii) lack of *language or institutional* support: first-generation migrant background (i.e., adolescent was born abroad). Adolescents for whom none of the three criteria applies are classified as "higher-SES". The categorization provides a roughly even split of our sample, with 46 percent classified as "low-SES".

While this multidimensional definition of SES aims to capture a broad set of dimensions of disadvantage, our results do not hinge on the specific categorization. In fact, we show that results are qualitatively similar for *each* of the underlying components (see section 6.3).[20] The idea behind the categorization is that each of the three dimensions of lacking family support leaves the adolescents considerably disadvantaged. The first dimension aims to capture low educational resources. This is partly captured by the fact that neither parent has attended university.[21] However, particularly in the German context with historically low university attendance, many

[19] The questionnaire item on the number of books at home contains six categories. We classify the lowest two categories – "none or only very few (0-10 books)" and "enough to fill one shelf" (11-25 books)" – as having few books, in contrast to the remaining four categories (everything above 25 books).

[20] Likewise, results hold for an SES index based on a principal component analysis of books at home, parental education, and parental employment (see Appendix E.3 of the working-paper version, Resnjanskij et al. (2021)).

[21] Many children do not know their parents' educational background. In our sample, 40 (32) percent of participants do not report the education level of the father (mother). Not being able to provide information on parental education is likely associated with low SES. For example, adolescents in the PISA 2012 sample who do not report the education for their parents rank on average at the 30th percentile in the PISA index for economic, social, and cultural status (ESCS) and also score very low in math performance (33rd percentile). Thus, our SES measure treats missing information on parental education as indicating non-university education.

parents without university education provide strong educational support. We capture the latter aspect by identifying only those parents without a university education who have few books in their home.[22] Second, families may lack the economic resources and time to support their children, which is particularly the case for single parents. To account for the fact that some well-off single parents are little constrained in supporting their children, we again combine this category with having few books at home. Third, first-generation migrants may additionally lack support not captured by the previous conditions, as they usually lack language skills, networks, experience, and institutional knowledge about the education system and the labor market.

A comparison of our sample to the general population of adolescents observed in the representative Programme for International Student Assessment (PISA) survey shows that respondents in our study are much more likely to come from disadvantaged households than the average adolescent in Germany.[23] Using our multidimensional SES measure, 46 percent of our sample are low-SES, compared to 24 percent in the adolescent population overall (see Appendix Table A1 for details). This greater disadvantage shows up in many dimensions, such as having few books at home (47 vs. 23 percent), living with a single parent (25 vs. 13 percent), first-generation migrant background (13 vs. 4 percent), and any migrant background (58 vs. 28 percent).[24] However, not everyone in our sample can be considered highly disadvantaged. There are pronounced differences between low-SES and higher-SES adolescents not only in the dimensions we used to define SES, but also beyond. For instance, 69 percent of fathers of low-SES adolescents are working full-time or part-time, compared to 81 percent of fathers of higher-SES respondents. Moreover, parents of low-SES respondents support their children significantly less with their homework than parents of higher-SES respondents (see Appendix Table A2). Notably, low-SES respondents also tend to be disadvantaged along several dimensions compared

[22] The number of books at home is a powerful proxy for the social, economic, and educational background of children's families (Schuetz, Ursprung, and Woessmann (2008)). Few books indicate low household possessions and thus low financial resources, but they also proxy for an otherwise low social and educational background. Intriguingly, in the international PIAAC data, the number of books at home in adolescence is strongly positively correlated with future earnings in *all* surveyed countries; on average across countries, respondents with more than 25 books at home at age 16 earn 22 percent more when aged 35-54 years (results available on request). Compared to other SES indicators such as parental education or income, books at home are also less prone to missing information. In our sample, all respondents provide information on the number of books in their home.

[23] The national PISA sample is representative of ninth-graders. We use the 2012 rather than 2015 PISA wave because it includes more variables that allow for a characterization of respondents' SES. The distribution of books at home is very similar in PISA 2012 and 2015, suggesting no discernible change in the SES of the pupil population.

[24] By contrast, the share without university-educated parents (69 vs. 70 percent) is not higher than in the overall population, which may partly reflect larger misreporting on this measure.

to similarly defined low-SES respondents in PISA, which partly reflects the substantially higher share of migrants in our sample.

At the same time, higher-SES adolescents in our sample cannot be considered as highly disadvantaged. For instance, the shares of university-educated fathers and mothers are higher in our higher-SES sample than in the representative PISA sample.[25]

### 4.2 Defining and Measuring Labor-Market Prospects

To measure the outcome of the mentoring treatment, we construct an index of labor-market prospects that combines three components: (1) school grades to measure a cognitive component; (2) patience and social skills to measure a behavioral component; and (3) labor-market orientation to measure a volitional component. The fact that participants in our evaluation are still attending school one year after program start, and a majority even three years after program start, precludes an extensive investigation of realized labor-market outcomes.[26] Therefore, in the pre-analysis plan, we defined the three outcome dimensions that are likely to be predictive of adolescents' long-term labor-market success.

We combine the three components in an overall index of labor-market prospects, but also report results for the three sub-indices. Apart from allowing for an overall assessment of program effectiveness, the index aggregation addresses concerns of multiple hypothesis testing by combining all outcome indicators into one measure and improves the statistical power to detect effects (Anderson (2008); Heller et al. (2017)). Following the procedure of Kling, Liebman, and Katz (2007), the overall index and all sub-indices that combine multiple survey items are each constructed as an equally weighted average of the standardized items that are included in the respective index. Each standardized item is a *z*-score itself, calculated by subtracting the control-group mean and dividing by the control-group standard deviation, separately by survey round.[27]

[25] For a comparison to the income of German households with a ninth-grade child overall, we can use entropy weighting to match our sample to the representative data of the German National Educational Panel Study (NEPS) based on five SES measures contained in both datasets. The approximated average household income of our higher-SES sample is just above the German mean, whereas the approximated average household income of our low-SES sample is only about two thirds of the national mean (detailed results available on request).

[26] In section 8.2, we report treatment effects on pursuing an apprenticeship three years after program start. Still, because the German education and training system offers many opaque preparatory options for graduates from low-track secondary schools whose effectiveness remains unclear for several years, measures allowing for a more encompassing evaluation of labor-market success will not be available until many years after the follow-up surveys.

[27] An index is computed for all individuals who have a valid response to at least one item; missing items for these individuals are imputed using the random-assignment group mean (see Kling, Liebman, and Katz (2007)).

### *a) Cognitive Component: School Grades*

***Relevance***. We measure the cognitive component of labor-market prospects by the math grades achieved in school. On the basis of representative skill assessment data from the Programme for the International Assessment of Adult Competencies (PIAAC) and PIAAC-L for Germany, we show that math grades in school are strong predictors of cognitive skills in adulthood (see Appendix E for the various PIAAC analyses mentioned here). Prior research suggests that cognitive skills, especially numeracy, are important determinants of individuals' wages and employment on the labor market, particularly in Germany (Hanushek et al. (2015)). We also find that better math grades in school are directly associated with higher wages and better employment opportunities. Conditional on math grades, German and foreign-language grades play little to no role for cognitive skills and labor-market success in adulthood. Moreover, migrants' school performance in language classes may suffer from measurement error.[28] Our analysis thus focuses on math grades.

***Measurement***. From the respective state administrative bodies, we obtained the permission to collect administrative data on school grades in math, German, and English directly from the schools. Data come from the pupils' report cards, which are issued after each school term (usually around February for the first half and around July for the second half of the school year) and are stored in the archives of local schools.[29] Grades are directly comparable between treatment and control pupils in each matched pair because the two pupils in each pair attend the same school and, in three quarters of cases, even the same classroom.[30] Grades are standardized and the usual German ordering is reversed so that higher values indicate better outcomes.

### *b) Behavioral Component: Patience and Social Skills*

***Relevance***. Our measure of the behavioral component of labor-market prospects combines patience and social skills. In line with the general literature on labor-market returns to non-

[28] Consistently, we do not find program effects on school grades in German or English (not shown).

[29] In cases where two parallel grading systems exist within a school that correspond to different school tracks, we use the official conversion tables provided by the respective state education ministry to convert all grades to the same grading system to ensure comparability within and across schools in a federal state. We also elicited grade information from the respondents. The correlation between administrative and self-reported math grades in the follow-up survey is high but not perfect ($r$=0.86), suggesting that the collection of administrative data reduced measurement error in the available grade information.

[30] In fact, dropping pairs in which treated and control respondents were not in the same classroom tends to increase estimated program effects on math grades (see section 6.2).

cognitive skills (e.g., Heckman, Stixrud, and Urzua (2006); Lindqvist and Vestman (2011)), there is increased attention to patience and social skills as predictors of labor-market success.[31]

Growing evidence suggests that higher levels of patience – as a measure of future orientation and willingness to postpone gratification – positively affect individuals' school achievement (Figlio et al. (2019); Castillo, Jordan, and Petrie (2019); Hanushek et al. (2022)) and labor-market success in adulthood (Golsteyn, Grönqvist, and Lindahl (2014)). Other concepts such as grit (defined as "perseverance and passion for long-term goals" (Duckworth et al. (2007)), conscientiousness, perseverance, and commitment, which are likely related to patience, have also been shown to be relevant for labor-market success (Almlund et al. (2011)). In our analysis of the German PIAAC data (which measures grit but not patience), higher levels of grit are associated with lower employment risk and higher wages particularly for low-SES individuals. In addition, higher levels of patience may increase the likelihood that adolescents continue and successfully complete an apprenticeship, particularly for low-SES individuals who are much more likely to quit an apprenticeship than higher-SES individuals.

Recent evidence also suggests growing importance of social skills and prosocial behavior on the labor market (e.g., Algan et al. (2016); Deming (2017); Kosse and Tincani (2020)). Another element of prosociality is trust (Kosse et al. (2020)) – i.e., beliefs about others' trustworthiness – for which evidence of relevance for labor-market outcomes is scarcer, with Butler, Giuliano, and Guiso (2016) as a noticeable exception. In the German PIAAC data, we find that trust is positively associated with employment prospects and wages.

***Measurement.*** We use survey responses to measure patience and social skills, relying on established taxonomies and survey items (see Appendix Table A3 for the underlying questionnaire items). The measure of patience is based on three survey items taken from the German Socio-Economic Panel (SOEP). The index of social skills comprises three sub-indices: prosociality, trust, and self-efficacy. Prosociality is measured by five items from the Strength and Difficulties Questionnaire (SDQ, see Goodman (1997)). Trust is measured by a survey item on general trust in people from the SOEP. Self-efficacy is measured by the three items of the General Self-efficacy Short Scale (Beierlein et al. (2012)).

[31] For example, the measure of non-cognitive ability used in Lindqvist and Vestman (2011) combines persistence, social skills, and emotional stability.

***c) Volitional Component: Labor-Market Orientation***

***Relevance***. An important aim of the mentoring program is to discover participants' potential and help them make up their minds about what they want to achieve in professional life. In Germany, the most promising career path for pupils in low-track schools, in particular for those with a non-academic family background, is to pursue an apprenticeship, which offers substantial returns on the labor market (e.g., Fersterer, Pischke, and Winter-Ebmer (2008); Piopiunik, Kugler, and Woessmann (2017)). Our PIAAC analysis shows that there is a large gap in the failure to obtain at least an apprenticeship-level professional qualification between low-SES individuals (21 percent) and higher-SES individuals (5 percent), which emerges already early in the career and is highly persistent. Moreover, university education does not seem to be a viable career path for the overwhelming majority of low-SES individuals, especially for those who attend low-track secondary schools (Appendix E.4). Therefore, the main goal of the mentoring program is to help disadvantaged participants in their transition into professional life by preparing them to find and successfully complete an apprenticeship.

***Measurement***. Our index of labor-market orientation combines two measures: the wish to conduct an apprenticeship and knowledge about the future career. The apprenticeship variable takes a value of one if respondents state that they would like to do an apprenticeship after school, and zero otherwise.[32] Career knowledge is measured by respondents' agreement to whether they already know exactly which occupation they want to work in later in life.

## 5. Empirical Strategy

This section shows that randomization led to balancing of our main variables between treatment and control groups (section 5.1) and introduces the estimation model (section 5.2).

### 5.1 Balancing of Baseline Characteristics

With the baseline survey administered before randomization, we can analyze the balancing of baseline variables in our sample. Columns 1-3 of Table 1 show that we do not observe meaningful pre-treatment differences between the treatment and control groups in any of the included baseline attributes. This indicates that the pair-wise matching procedure successfully generated balanced samples of treatment and control groups. Importantly, we also achieve

[32] The alternative answer categories are university, directly entering a job, other options, and not knowing yet.

balancing on variables not included in the matching approach: baseline values are balanced for all outcomes variables (panels A and B), for the variables used in the pair-wise matching (panel C), and for the control variables included in the main empirical specification (panel D).[33]

Since we investigate treatment effects separately for low-SES and higher-SES respondents, we also test for balancing by SES. To do so, we regress each baseline variable on the treatment indicator, a higher-SES dummy, and their interaction. Column 4 of Table 1 shows the $p$-value of an $F$-test of joint significance of the coefficients on the treatment indicator and its interaction with the higher-SES dummy. Results indicate that any differences between treatment and control groups in the baseline variables do not differ by SES. Thus, the randomization procedure achieves balancing in the full sample and in both SES subsamples.[34]

## 5.2 Estimation

Our empirical model is identified from the randomization of treatment. We define $Y_{ipt}$ as the post-treatment outcome of mentee $i$ in matching pair $p$ at time $t$ (i.e., about one, two, or three years after program start, depending on the survey). The treatment indicator $T_{ip}$ takes a value of one if the adolescent is offered to participate in the mentoring program, and zero otherwise. To test for heterogeneous treatment effects by SES, we interact the treatment indicator with an indicator for higher-SES background (from the baseline survey, period $t_0$), $HI\_SES_{ip(t_0)}$:

$$Y_{ipt} = \alpha_0 + \alpha_1 T_{ip} + \alpha_2 T_{ip} \times HI\_SES_{ip(t_0)} + \alpha_3 HI\_SES_{ip(t_0)} + \boldsymbol{X}'_{ip(t_0)}\boldsymbol{\alpha_4} + \mu_p + \epsilon_{ipt} \quad (1)$$

The vector $\boldsymbol{X}$ includes control variables from the baseline survey to improve precision of the estimation. Importantly, regressions control for the pre-treatment observation of the respective outcome variable. Additional pre-treatment control variables are gender, age, and migrant status as demographic variables; paid private teaching and parental homework support as non-

[33] Some baseline variables have a considerable number of missing values (Column 5 of Table 1). In particular, administrative math grades are missing for 88 respondents, either because they did not receive grades in the previous class (as is common in seventh grade in some schools) or because they changed schools before the current school year so that the current school could not provide the previous report card. Moreover, the question on the wish to get an apprenticeship after school is missing for 41 respondents because it was not part of the survey in the first pilot study. In order not to lose observations, we impute missing values of baseline variables with a constant and include missing indicators in all regressions. All index measures are based on non-imputed data only.

[34] Appendix Table A2 provides comparisons of the baseline variables between treatment and control groups within the subsamples of low-SES and higher-SES respondents, respectively. In the total of 90 comparisons across all three samples (full, low-SES, higher-SES), there is only one variable (self-efficacy in the higher-SES sample) that differs between treatment and control groups at a significance level of 5 percent and one (openness in the higher-SES sample) that differs at the 10 percent level.

mentoring-related types of school support; and the Big-5 personality traits (openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism) as description of adolescents' personality potentially relevant for labor-market prospects (see Appendix Table A3 for variable definitions). By virtue of our randomization approach, we can also include fixed effects $\mu_p$ for each matched pair. $\epsilon_{ipt}$ is an idiosyncratic error term.

The intention-to-treat effect (ITT) of being offered a place in the mentoring program for low-SES participants is given by $\alpha_1$. The coefficient $\alpha_2$ indicates how the treatment effect differs between higher-SES and low-SES participants. Since we sampled at the individual level, we provide robust standard errors, as well as $p$-values from a permutation test that randomly reassigns the treatment indicator within matched pairs (Heller et al. (2017); Abadie et al. (2020)).

We also estimate the treatment effect for adolescents who actually take up the program. Defining program take-up as the mentee having met the mentor at least once, we observe a take-up rate of 86 percent in the treatment group.[35] Take-up is somewhat lower for low-SES adolescents (82 percent) than for higher-SES adolescents (90 percent). We estimate the treatment effect on the treated (TOT) by two-stage least squares with random program assignment $(T_{ip})$ as an instrumental variable for actual participation $(D_{ip})$. The second stage (equation 3) uses participation $\widehat{D_{ip}}$ as predicted by assignment to treatment in the first stage (equation 2, as well as the prediction of $D_{ip} \widehat{\times HI\_SES}_{ip(t_0)}$ of the corresponding first stage):

$$D_{ip} = \gamma_0 + \gamma_1 T_{ip} + \gamma_2 T_{ip} \times HI\_SES_{ip(t_0)} + \gamma_3 HI\_SES_{ip(t_0)} + \boldsymbol{X}'_{ip(t_0)}\boldsymbol{\gamma_4} + \zeta_p + \omega_{ipt} \quad (2)$$

$$Y_{ipt} = \phi_0 + \phi_1 \widehat{D_{ip}} + \phi_2 \widehat{D_{ip} \times HI}\_SES_{ip(t_0)} + \phi_3 HI\_SES_{ip(t_0)} + \boldsymbol{X}'_{ip(t_0)}\boldsymbol{\phi_4} + \theta_p + \nu_{ipt} \quad (3)$$

Since we randomize at the individual level, we cannot rule out a priori that pupils in the control group benefit from treated peers, which would lead to an underestimation of the program impact. In the working-paper version (Resnjanskij et al. (2021), section 7.2), we provide a detailed analysis that the program is unlikely to have spillover effects on the control group, supporting the stable unit treatment value assumption (SUTVA) underlying the identification.

[35] The information about program take-up is based on mentee responses. In the few cases in which the mentee information is missing, we received information on the participation status from the mentoring sites.

## 6. Effects on Labor-Market Prospects at the End of the Program Period

This section reports effects of the mentoring program on various outcome measures at the end of the main program period, one year after program start, where we can still observe adolescents in the same school environment: the summary index of labor-market prospects (section 6.1), each of the three sub-indices that capture cognitive, behavioral, and volitional aspects of labor-market prospects (section 6.2), and further outcomes (section 6.3).

### 6.1 Index of Labor-Market Prospects

The index of labor-market prospects that combines math achievement, patience/social skills, and labor-market orientation provides an overall picture of the effectiveness of the mentoring program. Figure 2 shows treatment effects estimated in our baseline specification with all controls (see equation 1). The left panel indicates that program participation has a positive, albeit modest effect on average. One year after program start, the index of labor-market prospects for treated adolescents is 0.153 standard deviations higher than that of adolescents in the control group, significant at the 10 percent level ($p$=0.088).

The average effect masks considerable heterogeneity by SES background, however: highly disadvantaged participants benefit substantially from the program. The index of labor-market prospects for treated low-SES adolescents is 0.644 standard deviations higher than for low-SES adolescents in the control group ($p$<0.001; middle panel of Figure 2). This treatment effect even exceeds the SES gap in the index of labor-market prospects.[36] By contrast, the relatively more advantaged adolescents do not benefit from the program (right panel). If anything, they tend to be negatively affected. While the treatment effect of -0.221 standard deviations is economically relevant, it is not statistically significant at conventional levels.

Table 2 shows the corresponding regression estimates of the ITT effects for the full sample (panel A) and by SES (panel B). Treatment effects remain quite similar across specifications, but tend to become more precise once we control for the baseline survey information and matched-pair design. Unconditional treatment effects are shown in column 1. Column 2 adds the pre-treatment outcome as control, column 3 further includes a full set of fixed effects for the randomization pairs obtained from pair-wise matching, and column 4 additionally controls for

[36] The SES gap in labor-market prospects in the control group is 0.361 standard deviations (see Table 2, column 1 of panel B).

individual characteristics. Estimates of the average treatment effect in panel A are positive across specifications and become statistically significant when the pre-treatment outcome is controlled for. Panel B shows a large, highly significant treatment effect for low-SES adolescents. The large negative interaction between treatment and the higher-SES indicator shows that the treatment effect is significantly smaller for less deprived adolescents. The treatment effect for the higher-SES subgroup (reported at the bottom of the table) is negative, but not statistically significant.[37]

Estimates of the TOT effect of the mentoring program for those adolescents who actually took up the program are shown in column 5 of Table 2. With non-compliance, the TOT effect is larger than the ITT effect by the order of the inverse of the compliance rate. For low-SES adolescents, the TOT effect on the index of labor-market prospects is 0.771 standard deviations.

The effect of the mentoring program on low-SES adolescents is visible across the entire outcome distribution. With randomization, the distributions of treatment and control groups broadly overlap in the baseline survey for the full, low-SES, and higher-SES samples (Figure 3). One year after program start, the distribution of labor-market prospects of treated low-SES adolescents is strongly shifted to the right, with no substantive shift in the distribution of treated higher-SES adolescents. Analogously to the comparison of mean effects above, Kolmogorov-Smirnov tests reject the null hypothesis of equality of distributions between treatment and control groups in the full sample ($p$=0.083) and in the low-SES sample ($p$=0.002), but not in the higher-SES sample ($p$=0.750).

### 6.2 Sub-Indices of the Cognitive, Behavioral, and Volitional Components

In Table 3, we separate the index of labor-market prospects into its three sub-indices. For low-SES adolescents, the mentoring program has a significant positive effect on each of the three sub-indices – math achievement, patience/social skills, and labor-market orientation.[38]

[37] A sample split by SES background shows very similar (and even slightly stronger) treatment effects compared to the interaction specification (Appendix Table A4). Results are also similar (and again slightly stronger) when dropping pairs in which matched partners have a different SES. When exploring effect heterogeneity by gender, treatment effects are statistically significant for both low-SES males and low-SES females, with very similar point estimates (Appendix Table A5).

[38] The treatment effects on all three sub-indices remain statistically significant for low-SES adolescents when we correct for multiple hypothesis testing. We implement two corrections, the List, Shaikh, and Xu (2019) procedure based on Romano and Wolf (2010) and the Westfall and Young (1993) procedure, and the adjusted $p$-values range from 0.016 for both the patience and social skills index and labor-market orientation index using the Westfall-Young correction to 0.037 for math grades using the List-Shaikh-Xu correction (Appendix Table A6).

***Math Achievement in School.*** Column 1 of Table 3 reports program effects on administrative math grades in school. The outcome is *z*-standardized math grades (reversed order, such that higher values indicate better achievement). We find that participation in the mentoring program increases math achievement of low-SES adolescents by 0.309 standard deviations, closing almost three quarters of the SES achievement gap. The mentoring program does not significantly affect the math achievement of higher-SES adolescents. In three quarters of the matched pairs, both adolescents in the pair attend the same classroom. As grades are more comparable within a classroom, it is reassuring that the treatment effect on math grades tends to increase in the subsample of same-classroom pairs (column 2).

***Patience and Social Skills.*** Column 3 of Table 3 shows program effects on a summary measure that combines patience and social skills (equally-weighted average of *z*-scores of the two components). For low-SES adolescents, program participation increases the index of patience and social skills by 0.468 standard deviations, which is twice the SES gap in this outcome. The point estimate for higher-SES adolescents is insignificantly negative. Considering the separate components, treatment effects are more pronounced for patience than for social skills. Patience of low-SES adolescents responds strongly to the treatment (column 4).[39] The program effect of 0.444 standard deviations for low-SES adolescents exceeds the control-group gap in patience between higher-SES and low-SES adolescents (0.131 standard deviations). The treatment effect on the social-skills index is also positive for low-SES adolescents, but smaller (0.280 standard deviations) and not statistically significant at conventional levels (column 5). Treatment effects for all sub-indices of the social-skills index – prosociality, trust, and self-efficacy – are also positive for low-SES adolescents but never reach statistical significance.[40]

***Labor-Market Orientation.*** The mentoring program also raises the labor-market orientation of low-SES adolescents. Column 6 of Table 3 shows that treatment increases the index of labor-market orientation of low-SES adolescents by 0.443 standard deviations. Program effects on the labor-market orientation index of higher-SES adolescents are close to zero. Looking at the

[39] Mentoring may affect patience by taking over roles otherwise played by schools and parents in teaching forward-looking behavior and the ability to focus on the future (e.g., Becker and Mulligan (1997)). Recent examples show the malleability of patience by school-based interventions (Alan and Ertac (2018); Sutter et al. (2020)).

[40] Results available upon request. In line with the relatively weak effects on social skills, the mentoring program has no effect on social capital as measured by volunteering, the number of friends, and the frequency of meeting friends (Appendix Table A7). For low-SES adolescents, program participation neither affects a series of measures of school-related social capital. By contrast, for higher-SES adolescents the time spent in the program tends to crowd out school-related social activities, particularly low-stakes ones (see section 7 below).

separate components of the index, there is a sizeable treatment effect on highly disadvantaged youths' wish to get an apprenticeship after school.[41]

In addition, the program also increases adolescents' knowledge about their future career by 0.153 standard deviations (details not shown), which is, however, not statistically significant. We interpret this evidence that participants in the mentoring program get more realistic expectations about their future careers, as successfully completing an apprenticeship is the most relevant career track for disadvantaged youths in low-track schools (see section 4.2).

### 6.3 Further Analyses

Our main results are qualitatively similar when separating low-SES and higher-SES adolescents along the lines of each of the four components of the multidimensional SES measure. Table 4 shows results separately for the four components – books at home,[42] parental education, single-parent status, and first-generation migrant status. In each case, there are strong positive treatment effects for the low-SES subgroup, but not for the higher-SES subgroup. Effects are largest for the sub-group of first-generation migrants (0.753 standard deviations),[43] and only the single-parent dimension lacks statistical significance. Thus, results do not hinge on the specific definition of the multidimensional SES measure but emerge for the different dimensions of lack of family support (educational, economic/time, or language/institutional).[44]

---

[41] See Appendix Table A8 for details. There is some indication that the mentoring program provides potential-specific career guidance, as there is a positive (albeit insignificant) treatment effect on the wish to study at university for higher-SES adolescents. Moreover, the wish to get an apprenticeship is sizably (but statistically insignificantly) stronger in the subsample of adolescents with lower grades in the pre-treatment period.

[42] We asked adolescents in the survey three years after program start once again about the number of books at home when they were 13 years old. The correlation with the measure three years earlier in the baseline survey is 0.78, indicating high reliability. When we use the later books-at-home information to classify respondents as low-SES, the treatment effect for low-SES adolescents is 0.518 (s.e. 0.180) compared to the 0.556 shown in Table 4. We also find similar qualitative results when classifying adolescents as low-SES when they report few books either in both waves or in at least one of the two waves. In line with attenuation bias in the SES classification, the treatment effect estimate increases to 0.652 (s.e. 0.259) when we instrument the initial observation of books at home (and its interaction with the treatment indicator) by the later observation (and its interaction with the treatment indicator).

[43] Treatment effects are also significantly positive, at 0.290 standard deviations, for the overall (first- and second-generation) migrant subpopulation (Appendix Table A9), which constitutes 58 percent of our sample. However, effects for the subgroup of second-generation migrants (adolescents who have at least one parent born abroad, but who are born in Germany themselves) are only weakly positive and not statistically significant.

[44] The interpretation that the mentoring is particularly effective in settings where adult support is lacking is also consistent with the finding that treatment effects are substantially larger for adolescents who do not receive any homework support by their parents (at baseline) than for those who receive strong homework support (available on request). However, this effect heterogeneity is dominated by the heterogeneity along the SES dimension shown here.

As measures of overall program effectiveness, we also investigate various dimensions of satisfaction (Table 5). For low-SES adolescents, participating in the mentoring program leads to higher life satisfaction: low-SES youths in the treatment group are 24.6 percent more likely to be satisfied with their lives than their counterparts in the control group (column 1). They are also more satisfied with their current belongings (column 2), suggesting that the program makes highly disadvantaged adolescents focus on what they can realistically achieve and appreciate what they already possess. Furthermore, there are (insignificant) positive effects on low-SES adolescents' satisfaction with their current school situation (column 3) and their math performance (column 4); the fact that the latter effect is weaker than the effect on actual math performance may indicate that treated adolescents may have had even higher aspirations. There are no significant treatment effects on satisfaction for higher-SES youths.

## 7. Analysis of Mechanisms

This section studies a range of potential channels that might underlie the treatment effect of the mentoring program. Our mediation analysis follows the approach developed in Heckman, Pinto, and Savelyev (2013) and Heckman and Pinto (2015), which provides a decomposition of the overall treatment effect into shares attributed to different mediators.[45] Because of the opposing effects, we implement the mediation analysis separately for low-SES and higher-SES adolescents. Thus, our baseline equation 1 simplifies to a regression of the outcome $Y_{it}$ on the treatment indicator $T_i$ and baseline covariates $\boldsymbol{X}_{i(t_0)}$ in the respective subsample:[46]

$$Y_{it} = \beta_0 + \beta_1 T_i + \boldsymbol{X}'_{i(t_0)}\boldsymbol{\beta_2} + \varphi_{it} \qquad (4)$$

Our main focus is to analyze the positive program effect for low-SES youths. As potential mediators, we consider several aspects of the mentor-mentee relationship that are potentially related to developing a career vision for low-SES adolescents and facilitating their transition into professional life. Since the one-to-one mentoring is at the core of the mentoring program, we

[45] See, e.g., Oreopoulos, Brown, and Lavecchia (2017) and Kosse et al. (2020) for applications. While inclusion of the potential mediator variables in the questionnaire indicates that we planned their analysis, we did not specify any of the specifics of the mediation analysis in advance. Therefore, this section is part of the exploratory data analysis that mainly aims to inform future research that digs deeper into which specific aspects of mentoring programs are key to success.

[46] As the sample split sometimes cuts through pairs with different SES and the subsample results are very similar with and without randomization-pair fixed effects (Appendix Table A4), we do not use randomization-pair fixed effects in these specifications.

expect that the program's success hinges on whether or not the mentors provide adult support for future-related issues, which the disadvantaged adolescents potentially lack. In particular, we focus on three potential mediating factors – elicited for both treatment and control groups in the background questionnaires – that proxy for mechanisms that are each related to one of the three components of labor-market prospects that we consider in our baseline analysis: schools, future orientation, and occupational orientation.

The first mediator captures whether, as part of developing a career vision, the mentoring is successful in making mentees perceive school as useful for a later job. The second mediator reflects to which extent the treatment can successfully establish the mentor as an attachment figure for talking about the future. The third mediator captures whether mentors are important for providing information about occupational choice. Program participation indeed positively affects all three mediators for low-SES adolescents – the extent to which they agree that material learnt in school is useful for future jobs, the likelihood that they mention a mentor or coach as a person with whom they talk about their future, and the likelihood that they consider a mentor or coach as an important source of information for job choice – and all three mediators are significantly related to the outcome index for low-SES adolescents (see Appendix F for details).

The mediation approach assumes that the outcome can be expressed as a linear combination of the $k = 3$ mediators $M_{it}^k$ and a vector of baseline demographic characteristics $\boldsymbol{X}_{i(t_0)}$. This allows us to rewrite equation 4 as:

$$Y_{it} = \beta_0 + \beta_1^{residual} T_i + \sum_k \theta^k M_{it}^k + \boldsymbol{X}'_{i(t_0)} \boldsymbol{\beta_2} + \mu_{it} \tag{5}$$

The coefficient $\beta_1^{residual}$ represents the effect of the mentoring program that is not explained by changes in the observed mediators. Consequently, the share of the treatment effect that is explained by the combined changes in the observed mediators is given by $1 - \beta_1^{residual}/\beta_1$.

Assessing the *relative* contribution of the different mediators additionally requires estimates of the effects of the treatment on the respective mediators:

$$M_{it}^k = \delta_0^k + \delta_1^k T_i + \boldsymbol{X}'_{i(t_0)} \boldsymbol{\delta_2^k} + \nu_{it} \tag{6}$$

The share of the overall treatment effect that can be attributed to the $k^{\text{th}}$ mediator can then be calculated by multiplying the treatment effect on the mediator $\delta_1^k$ with the impact of the mediator on the outcome $\theta^k$ and dividing by the reduced-form treatment effect on the outcome $\beta_1$:

$$share\ M_k = \theta^k \delta_1^k / \beta_1 \tag{7}$$

By the virtue of randomization $\beta_1$ and $\delta_1^k$ are identified, whereas the analysis has to rely on the arguably strong assumption of conditional independence to identify $\theta^k$.[47] The assumption implies that any potential unobserved mediator or control subsumed in the error term $\mu_{it}$ is orthogonal to the included mediators. For example, if (a) treatment affects both job-choice information from the mentor (included in our mediation analysis) and connections to potential training firms (which we do not observe), (b) these two are relevant for our outcomes, and (c) they are positively correlated, the estimates of $\theta^k$ for the mediator "mentor important for job choice" will be upward biased. As a consequence, the estimated shares of the treatment effect attributed to the mediators should be interpreted as upper bounds.

Figure 4 shows the results of the mediation analysis that considers the three mediators in explaining the effect of the mentoring program for low-SES adolescents (see Appendix F.1 for details). Focusing on the overall index of labor-market prospects as the outcome, panel A decomposes the overall treatment effect into shares attributed to changes in the three mediator variables. Considered separately in the first three bars, changes in perceiving school as useful for later jobs account for 7 percent of the overall treatment effect, talking with the mentor about the future for 31 percent, and considering the mentor as an important source of information for job choice for 16 percent. Considering the three mediators jointly in the fourth bar indicates that the latter effect mostly materializes through talking with the mentor about the future. Together, the three mediator variables account for 37 percent of the overall treatment effect, with the bulk attributed to whether the mentor acts as an attachment figure to whom the low-SES adolescents talk about their future. Given the proxy nature of the mediator variables, this is a substantial attribution that provides relevant hints on underlying mechanisms; at the same time, the majority of the overall treatment effect cannot be accounted for by the observed mediator variables.

Panel B of Figure 4 provides equivalent decompositions for each of the three components of the index of labor-market prospects. The combined mediators account for between 29 and 60 percent of the treatment effects on the separate components. Interestingly, talking with the mentor about the future is mainly responsible for the treatment effect on math achievement, whereas somewhat surprisingly, an increased perception of school as useful for jobs does not

[47] To ensure that the estimate of $\theta^k$ is not affected by preexisting differences in the baseline covariates or mediators, we additionally control for the baseline values of the SES-specific mediator variables when available.

mediate this effect. Talking with the mentor about the future also accounts for most of the treatment effect on patience and social skills. This is consistent with the idea that talking about future-related issues raises the awareness of the importance of current investments (in education, job applications, social behavior, etc.) that may pay off later in life (e.g., in terms of better labor-market outcomes). The treatment effect on labor-market orientation is largely driven by mentors' guidance concerning potential future jobs.

Expectedly, the set of mediators considered in the low-SES analysis does a poor job in explaining the negative higher-SES treatment effect. The treatment does affect two of the three mediators – talk to mentor about future and mentor important for job choice – also for higher-SES adolescents. However, among higher-SES adolescents, these mediators are not significantly associated with labor-market prospects (detailed results available on request). This supports the interpretation that higher-SES adolescents do not lack these kinds of resources, so that the mentoring program cannot substitute a lacking resource.

Instead, conducting a mediation analysis for the higher-SES adolescents indicates that their (insignificant) negative treatment effect can partly be attributed to a crowding-out of in-school social activities and performance appreciation (see Appendix F.2 for details). The time that mentees spend with the mentors may in principle crowd out participation in other useful activities. Indeed, for higher-SES (but not for low-SES) adolescents we find that the mentoring program leads to significant reductions in school-related social activities and in the perceived importance of good grades. Together, these two factors account for about 60 percent of the negative higher-SES treatment effect in a mediation analysis, with the crowding-out of social activities in school as the dominant channel.

Finally, a descriptive analysis does not show extensive differences in characteristics of the mentoring relationship that could account for why the mentoring program affects low-SES but not higher-SES adolescents. Information on the mentoring relationships elicited from the adolescents in the treatment group one year after program start indicates that the initiation and continuation of the relationships, as well as the frequency, length, and content of the meetings, does not differ significantly between low-SES and higher-SES mentees (see Appendix Table A10). The one exception is that higher-SES mentees are more likely than low-SES mentees to talk with their mentors about leisure activities (63 vs. 49 percent). While more low-SES than higher-SES mentees think that their school performance increased due to their mentors (26 vs. 16

percent) and that their mentor was helpful in tackling problems outside school (36 vs. 26 percent), these differences do not reach statistical significance at conventional levels. Thus, differences in the characteristics of the mentoring relationships do not seem to contribute to the differences in treatment effects by SES.

## 8. Persistence and Early Transitions into Professional Life

With our additional follow-up surveys conducted two and three years after program start, we can test whether effects on the cognitive, behavioral, and volitional components of labor-market prospects persist beyond the program period (section 8.1) and whether the program affected early realizations of school-to-work transitions (section 8.2).

### 8.1 Persistence of Effects on Components of Labor-Market Prospects

The one-year effects presented so far indicate important effects of the mentoring program by the end of the main program period. By observing the different components of labor-market prospects two and three years after program start, we can study whether treatment effects persist after the program has ended.[48] However, this analysis is somewhat limited by the fact that by this time, many adolescents have left their initial classes for newly formed classes, different types of schools, or even the labor market or unemployment. These changing contexts require various adjustments to the measurement of outcomes to allow for consistent comparison over time.[49]

Results indicate that large treatment effects on the math performance of low-SES adolescents persist to the end of low-track secondary school and that the treatment effects on labor-market orientation even increase until three years after program start, whereas treatment effects on patience and social skills fade away after the mentoring program has ended. Consistent with the one-year effects, there are no significant longer-term effects on either component of labor-market prospects for higher-SES adolescents.

***Math Achievement in School.*** Collection of subsequent administrative data on math grades allows us to observe adolescents' school performance until the end of low-track secondary

[48] In the second follow-up survey, we observe that 37 percent of mentees still had a mentor in the second year.

[49] Consistent measurement of outcomes may be further affected by the Covid-19 situation, which impacted some observations at the end of the survey period two years after program start and most observations in the survey period three years after program start.

school, which corresponds to two years after program start.[50] To ensure comparability of math grades over time, we limit the persistence analysis to those 63 percent of randomized pairs whose two adolescents still attend the same class two years after program start. Estimating treatment effects in the full sample would likely yield downward biased estimates for two reasons. First, the limited comparability of grades across teachers and classrooms induces measurement error in grades. Second, because the treatment tends to increase the likelihood of graduating in a higher academic track (see below), treated adolescents are exposed to better-performing peer groups, rendering it more difficult to achieve good grades if there is grading on a curve.

Participating in the mentoring program increases low-SES adolescents' final math grade in low-track secondary school, two years after program start, by 0.353 standard deviations (panel A of Figure 5). This is somewhat lower than the estimate in the first period, but remarkably stable over the subsequent two periods.[51]

We also collected administrative information on graduation outcomes. There is a positive, but rather small and insignificant treatment effect on the probability of low-SES adolescents to graduate from low-track secondary school, which mostly reflects ceiling effects as 89 percent of the control group have graduated by the end of the observation period. Pupils from low-track schools can graduate either with a basic or an upper certificate (*Erster allgemeiner* vs. *Mittlerer Schulabschluss*). Intriguingly, the program has an economically meaningful effect on graduating with an upper certificate. While imprecisely estimated and not statistically significant at conventional levels, the treatment effect for low-SES adolescents is 8.6 percentage points, or about 15 percent relative to the control-group mean (see Appendix Table A11 for details).

***Patience and Social Skills.*** Results do not suggest persistent effects on patience and social skills two and three years after program start. The positive treatment effect on low-SES adolescents' patience after the first year fades out completely by the second year after program start (panel B of Figure 5). Similarly, there are no treatment effects on measures of social skills.[52]

[50] The additional data points for which administrative grades can be observed effectively refer to the report cards obtained three quarters of a year and one and a half years, respectively, after program start.

[51] See Appendix Table A11 for the corresponding regression estimates. Consistent with downward bias in the full sample, treatment effects on math grades are only about one-third of the effect size in the full sample and not statistically significant (results not shown).

[52] Since there were no significant treatment effects on social skills (i.e., prosociality, trust, and self-efficacy) one year after program start, we collected only limited social-skill information in the subsequent waves – social skills two years after program start and trust three years after program start. Consistent with the initial results, none of the social-skill measures is significantly affected by the treatment in the following years (results not shown).

The result on patience suggests that youths' future orientation increases when they talk with their mentor about future-related topics, but falls back to the level of the control group when the relationship has ended. This short-lived effect on patience does not necessarily imply that the initial effect on patience did not have longer-term repercussions, though, as it may have triggered relevant decisions at the time that persistently affect other relevant life outcomes. In particular, the increase in patience occurred at a time when important decisions about the transition to the labor market – e.g., whether or not to search for and pursue an apprenticeship – had to be made.

***Labor-Market Orientation.*** The treatment effect on the labor-market orientation of low-SES adolescents, already substantial one year after program start, persists in the second year and increases further three years after program start (panel C of Figure 5). At that time, the labor-market orientation index of treated low-SES adolescents is 0.847 standard deviations higher.[53]

Three years after program start, we also elicited information about the occupation that adolescents want to get in the future. Overall, 69 percent of respondents can name at least one desired occupation. Consistent with the positive treatment effects on the index of labor-market orientation, treated low-SES adolescents are a significant 18.6 percentage points (27 percent of the control-group mean) more likely to report a desired occupation (Appendix Table A12). Defining the outcome as either reporting a desired occupation or already having an occupation, the treatment effect even increases to 26.4 percentage points (38 percent of the control-group mean). Consistent with these results, adolescents in the treatment group consider their mentor as helpful for their career choice (not shown). These findings confirm that the mentoring program improved labor-market orientation by building up a vision of the adolescents' future career.

### 8.2 Early Realizations of Labor-Market Transitions

The primary goal of the mentoring program is to improve transitions from school into professional life. Even three years after program start, the majority of adolescents (56 percent) is still in school. However, some mentees have transitioned into the labor market by that time, so

[53] In the labor-market orientation index, the question regarding knowledge about the future career remained the same over all survey waves. By contrast, we had to adapt the question regarding apprenticeships in later surveys to capture that adolescents already had the opportunity to actually enter the labor market. Thus, we replaced the question asking for the adolescents' wish to pursue an apprenticeship after school with a more detailed question about specific plans after finishing school (second follow-up survey) and future aspirations with respect to the educational degree (third follow-up survey, see Appendix Table A3). In both surveys, we collapse answers to a dummy variable taking a value of one if the adolescent wants to pursue an apprenticeship, and zero otherwise. While this adaptation somewhat limits the comparability of results over time, it is still useful to get an idea about the evolution of the adolescents' labor-market orientation.

we can observe early career realizations for some of the adolescents. Importantly, a substantial share of adolescents had the opportunity to start an apprenticeship, which is a key outcome for low-track school graduates in Germany (see section 4.2) and an explicit goal of the program. Thus, while ultimate labor-market outcomes will not be realized for many years to come, these first school-to-work transitions allow for an early look into emerging labor-market patterns.

Results show that the program has a large positive effect on low-SES adolescents' probability to pursue an apprenticeship three years after program start. The treatment effect of 29.3 percentage points corresponds to a doubling of the share of adolescents doing an apprenticeship compared to the control group (column 1 of Table 6). Consistent with the strong and persistent effect on labor-market orientation shown above, these results indicate that the mentoring program significantly affects the transition of low-SES adolescents into the labor market. For higher-SES adolescents, the treatment effect on pursuing an apprenticeship is negative and not statistically significant.

The most likely counterfactual early-career outcomes of treated low-SES adolescents would have been further attending school (partly in a preparatory system with unclear effectiveness) or suffering from unemployment. Although the estimation lacks the power to identify treatment effects precisely, the mentoring program has sizeable negative effects on attending school of 20.5 percentage points and on being unemployed or pursuing other non-school or non-work-related activities by 11.7 percentage points (columns 2 and 4 of Table 6). By contrast, the treatment does not affect conducting work-related activities such as pre-employment training, internships, or a voluntary social year (column 3).

One potential mechanism explaining the treatment effect on early school-to-work transitions is that treated adolescents have more realistic career expectations, in particular, regarding the probability to finish university studies. In the control group, 63 percent of low-SES adolescents believe that they can successfully complete university studies, while the actual share of low-SES individuals from low-track secondary schools with a university degree is just 2 percent in Germany (see Appendix E.4). Treatment reduces low-SES adolescents' self-assessed likelihood to successfully complete university education by 13.4 percentage points (Appendix Table A13), suggesting that treated adolescents have more accurate beliefs regarding the likelihood of graduating from university. One interpretation of this result is that mentors, who are university students themselves, can provide meaningful advice on what it takes to successfully complete a university education. At the same time, the mentoring program affects neither the self-assessed likelihood of finishing apprenticeship training nor the expected earnings returns of completing either university studies or apprenticeship training (Appendix Table A13).

The more realistic expectations regarding career opportunities in the treatment group do not negatively affect satisfaction with the current situation; on the contrary, the treated low-SES youths consider their current career path more desirable. They are 31.2 percentage points more likely to be satisfied with the current situation than the control group (column 5 of Table 6). They are also 22 percentage points more likely to not want to change their current situation (column 6). By contrast, satisfaction levels in the higher-SES sample are only modestly and insignificantly affected by the treatment.

Calculating the earnings to be expected from adolescents' current or desired occupations suggests that the mentoring program does not nudge adolescents into lowering their career ambitions. A possible concern with the satisfaction results could be that they are driven by a lower career ambition of treated adolescents. It is also conceivable that they are not aware that better career outcomes could potentially be reached by continuing formal education. To address these issues, we extrapolate expected earnings based either on adolescents' actual apprenticeship occupation or on their desired occupation elicited three years after program start.[54] We use administrative data from the Federal Employment Agency to infer the earnings that adolescents can expect to receive if they worked in the reported (five-digit) occupation, considering the occupation median in monthly earnings in 2020. Expected earnings of treated low-SES adolescents are in fact *higher* than those in the control group. For instance, low-SES adolescents in the treatment group can expect monthly earnings of 3,066 EUR if they continued to work in their current apprenticeship occupation, compared to 2,746 EUR in the control group. This gap decreases only modestly when considering the desired occupation (3,406 EUR in the treatment group vs. 3,184 EUR in the control group). Thus, treated low-SES adolescents have the ambition to get into better-paying jobs than those in the control group. Treated adolescents also aim at working in jobs that are less substitutable by technology (e.g., Autor (2022)): the automation

[54] We elicited the desired occupation by first asking respondents: "Do you already know which occupation you want to take up?" Answer categories were: yes, with great certainty; yes, with some certainty; no, still open. Only respondents who answered yes (i.e., the first two categories) were asked about the specific occupation in which they would like to work. We assigned respondents' free text answers to the German five-digit occupational classification (KldB). 85 adolescents could not provide a desired occupation. Results when imputing missing values with the treatment-specific mean are qualitatively similar to those without imputation reported here.

probability of the desired jobs of low-SES adolescents in the treatment group is 37.1 percent, compared to 43.3 percent in the control group.[55] Together, the results suggest positive effects of the mentoring program on low-SES adolescents' emerging transitions into the labor market.

## 9. Cost-Benefit Analysis and Program Scalability

In this section, we address two important questions regarding the policy implications of our results: a cost-benefit analysis (section 9.1) and the potential reach of the program (section 9.2).

### 9.1 Cost-Benefit Analysis

The cost-benefit analysis suggests that the mentoring program is highly cost-effective. We quantify benefits by the expected lifetime labor-market returns from improved school performance due to program participation, accounting for the fade-out in treatment effects on math grades apparent in Figure 5. Given the large program effect, the projected gain in discounted lifetime earnings amounts to 13,500 EUR for low-SES adolescents (see Appendix G for details). By contrast, actual program costs are relatively low at 750 EUR per participant. The program thus yields benefit-cost ratios that range from 8-to-1 for an untargeted program to 18-to-1 for a program targeted at low-SES adolescents – a similar ballpark to, e.g., the crime-reduction intervention studied by Heller et al. (2017). Although the cost-benefit analysis should be regarded as back-of-the-envelope calculation with considerable uncertainty, the large magnitude of the estimates suggests that the costs of the mentoring program are likely more than offset by the long-term earnings benefits it generates.

### 9.2 Program Scalability

The positive cost-benefit assessment begs the question of scalability of successful mentoring programs. In this section, we show that it is possible to screen applicants who are most likely to benefit from the program, that the program effects are scalable beyond one specific site, that the reach of the program could be substantially expanded, and that the funding would only require a relatively small share of the cities' budgets for social expenditures.

First, the strong heterogeneity of results by SES suggests that to have impact, a scaled program should be targeted at those youths who really lack family support. Other adolescents

[55] Automation probability is defined as the share of tasks in an occupation that can be replaced by technology. Data are provided by the Institute for Employment Research (IAB, see https://job-futuromat.iab.de/en/).

with a more favorable family environment, even if disadvantaged in other regards, do not seem to benefit from the program. The positive aspect of this is that, almost by definition, the low-SES subgroup is the main target group for policies that aim to reduce persistence in inequality by spurring upward intergenerational mobility. In practical terms, administrators would have to screen potential participants before program start and admit only those youths for whom the program is likely effective.

In practice, screening could be performed based on various SES indicators. The fact that there are strong treatment effects for low-SES adolescents both for the multidimensional SES measure and for each of its four components (Table 4) suggests that the program is likely to be effective regardless of the specific criteria used to classify low-SES adolescents in order to screen applicants. Furthermore, while it may not be practically feasible to screen applicants based on the number of books at home, it would be straightforward to screen applicants based on their parents' education, their single-parent status, or their first-generation migrant status.

Second, the design of the field experiment was geared to show scalability beyond one specific location because it was not restricted to one or two selected sites, but administered in 19 schools in nine mentoring sites across Germany, ensuring that treatment effects are not driven by any specific location. Estimated site-specific treatment effects for low-SES adolescents are positive for *each* individual site (see Appendix Figure A1). In seven of the nine sites, the estimated point estimate is larger than the average effect found in our baseline model, and (despite the small site-specific sample sizes) it is separately statistically significant in six of them. Furthermore, the size of the estimated treatment effect does not vary systematically with city size. The consistency of the site-specific estimates speaks in favor of general scalability.

Third, the franchise has grown from one to over forty locations within just ten years, showing its growth potential. As a limiting factor, the program relies on university students as mentors and thus only runs in cities with universities, so the evaluation cannot speak toward generalizability to rural areas without higher-education institutions. In fact, though, more than two thirds (69 percent) of adolescents who attend low-track secondary schools in Germany live in cities/counties that have a university. Also more generally, the most disadvantaged adolescents typically live in larger cities, such as Berlin and Hamburg, instead of smaller towns or rural areas without a university. To quantify the potential reach of a fully-scaled targeted mentoring program, we calculate the number of disadvantaged adolescents who attend low-track secondary schools in university cities/counties in school year 2017/18, the starting year of our RCT (see Appendix H.1 for details). According to our estimates, the potential reach of the program is about 134,000 adolescents each year, or 21 percent of an entire cohort (comprising pupils from all school tracks) in Germany. Even when restricting the analysis only to the cities in which the mentoring program is already active, the potential reach of the program is over 36,000 adolescents. Only 2 percent of these potential mentees are currently covered by the studied mentoring program, suggesting substantial scope for scaling up the program.

However, scalability may be restricted by the availability of volunteer mentors. Assuming that mentors are only recruited from the pool of freshmen students and applying the share of 15 percent of current students who engage in voluntary activities in areas related to youth and social work, we estimate the total potential supply of mentors at roughly 75,000 each year. This would mean that 56 percent of the potential reach could be served (12 percent of a cohort). This supply restriction may or may not be binding, depending on the share of the adolescents in the potential reach who are willing to participate in the program.

Fourth, sustained funding is an important prerequisite for scaling any program. Importantly, the program is not very expensive because it draws on volunteer students. Even when adding opportunity costs of the mentors to the cost-benefit analysis indicated above, each mentoring relationship costs about 1,950 EUR. A fully-scaled targeted program would thus require a budget of about 145 million EUR per year. While this is reasonably large for Germany overall, it boils down to just 1-3 percent of the household budgets for social expenditures for cities such as Aachen and Cologne (see Appendix H.2 for details). It thus seems straightforward that local city governments could provide sustained financing for a fully-scaled mentoring program.

## 10. Conclusion

Our results suggest that mentoring programs can successfully improve the future labor-market opportunities of highly disadvantaged youths. At the end of the main program period, one year after program start, the mentoring program that we study increases a summary measure of labor-market prospects for low-SES adolescents by more than half a standard deviation, fully closing the SES gap. All three components of the summary measure – capturing cognitive, behavioral, and volitional aspects – are positively affected. The positive effects on the cognitive and volitional components, but not the behavioral component, persist three years after program start, well after the end of the program. By that time, positive effects are also visible on early realizations of labor-market transitions, i.e., low-SES adolescents' likelihood to pursue an apprenticeship.

Therefore, mentoring seems a viable policy to raise the prospects of disadvantaged children even at adolescent age. Of course, mentors can never fully substitute for parents, and they never aim to. However, by providing guidance for future opportunities, they appear to be able to substitute for some elements of parental support that many disadvantaged youths are lacking. Part of the overall program effect can be accounted for by aspects of the mentor-mentee relationship that help low-SES adolescents develop a career vision, in particular guidance by the mentors for their future.

By contrast, the program does not significantly affect higher-SES adolescents, and if anything, estimates point in the negative direction. Lack of adult support does not seem to be a major handicap for these relatively less disadvantaged youths. Compared to low-SES participants, they are less likely to consider their mentors as a helpful resource for solving problems inside and outside of school, and program participation may even crowd out their participation in social school activities. To the extent that the counterfactual time use in any crowded-out activities in families, schools, or elsewhere would have been more beneficial for their future labor-market opportunities, the program might even be harmful to higher-SES adolescents. On the other hand, these adolescents might also derive other benefits from program participation that are not captured by our outcome measures, such as gaining new perspectives, building additional network relations, or pure consumption value.[56]

A key policy implication is that to be successful, mentoring programs should be targeted to adolescents from disadvantaged families. Cost-benefit considerations suggest that targeted programs can be highly cost-effective. Furthermore, scalability considerations suggest that the program's potential reach is quite large.

[56] A particular consumption value of the mentoring relationship for higher-SES adolescents would be consistent with the finding that the only dimension in which the content of their mentoring meetings differs from that of low-SES adolescents is that they are more likely to talk with their mentor about leisure activities. Still, we do not know the consumption value of the relevant counterfactual activity.

## References


Abadie, Alberto, Susan Athey, Guido W. Imbens, Jeffrey M. Wooldridge (2020). Sampling-Based Versus Design-Based Uncertainty in Regression Analysis. *Econometrica* 88 (1): 265-296.

Alan, Sule, Seda Ertac (2018). Fostering Patience in the Classroom: Results from Randomized Educational Intervention. *Journal of Political Economy* 126 (5): 1865-1911.

Algan, Yann, Elizabeth Beasley, Richard E. Tremblay, Frank Vitaro (2016). *The Impact of Non-Cognitive Skills Training on Academic and Non-Academic Trajectories: From Childhood to Early Adulthood*. http://www.yann-algan.com/en/research-articles/impact-non-cognitive-skills-training-academic-non-academic-trajectories-childhood-early-adulthood-2/ (Accessed September 4, 2020).

Algan, Yann, Christian Dustmann, Albrecht Glitz, Alan Manning (2010). The Economic Situation of First and Second-Generation Immigrants in France, Germany and the United Kingdom. *Economic Journal* 120 (542): F4-F30.

Almlund, Mathilde, Angela L. Duckworth, James Heckman, Tim Kautz (2011). Personality Psychology and Economics. In *Handbook of the Economics of Education, Vol. 4*, edited by Eric A. Hanushek, Stephen Machin, Ludger Woessmann. Amsterdam: North Holland: 1-181.

Almond, Douglas, Janet Currie, Valentina Duque (2018). Childhood Circumstances and Adult Outcomes: Act II. *Journal of Economic Literature* 56 (4): 1360-1446.

Alvaredo, Facundo, Lucas Chanel, Thomas Piketty, Emmanuel Saez, Gabriel Zucman, eds. (2018). *World Inequality Report 2018*. Paris: World Inequality Lab.

Anderson, Michael L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevalution of the Abecedarian, Perry Preschool, and Tearly Training Projects. *Journal of the American Statistical Association* 103 (484): 1481-1495.

Autor, David H. (2014). Skills, Education, and the Rise of Earnings Inequality among the "Other 99 Percent". *Science* 344 (843): 843-851.

Autor, David H. (2022). The Labor Market Impacts of Technological Change: From Unbridled Enthusiasm to Qualified Optimism to Vast Uncertainty. NBER Working Paper 30074. Cambridge, MA: National Bureau of Economic Research.

Becker, Gary S., Casey B. Mulligan (1997). The Endogenous Determination of Time Preference. *Quarterly Journal of Economics* 112 (3): 729-758.

Beierlein, Constanze, Anastassyia Kovaleva, Christoph J. Kemper, Beatrice Rammstedt (2012). Ein Messinstrument zur Erfassung subjektiver Kompetenzerwartungen: Allgemeine Selbstwirksamkeit Kurzskala (ASKU). GESIS-Working Paper 2012|17. Mannheim: GESIS – Leibniz-Institut für Sozialwissenschaften.

Black, Sandra E., Paul J. Devereux (2011). Recent Developments in Intergenerational Mobility. In *Handbook of Labor Economics, Vol. 4, Part B*, edited by Orley Ashenfelter, David Card. Amsterdam: North Holland: 1487-1541.

Borghans, Lex, Bart H. H. Golsteyn, James J. Heckman, John Eric Humphries (2016). What Grades and Achievement Tests Measure. *Proceedings of the National Academy of Sciences* 113 (47): 13354-13359.

Bruhn, Miriam, David McKenzie (2009). In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics* 1 (4): 200-232.

Butler, Jeffrey V., Paola Giuliano, Luigi Guiso (2016). The Right Amount of Trust. *Journal of the European Economic Association* 14 (5): 1155-1180.

Castillo, Marco, Jeffrey L. Jordan, Ragan Petrie (2019). Discount Rates of Children and High School Graduation. *Economic Journal* 129 (619): 1153-1181.

Corak, Miles (2013). Income Inequality, Equality of Opportunity, and Intergenerational Mobility. *Journal of Economic Perspectives* 27 (3): 79-102.

Cunha, Flavio, James J. Heckman, Lance Lochner, Dimitriy V. Masterov (2006). Interpreting the Evidence on Life Cycle Skill Formation. In *Handbook of the Economics of Education, Vol. 1,* edited by Eric A. Hanushek, Finis Welch. Amsterdam: North Holland: 697-812.

Deming, David J. (2017). The Growing Importance of Social Skills in the Labor Market. *Quarterly Journal of Economics* 132 (4): 1593-1640.

DuBois, David L., Bruce E. Holloway, Jeffrey C. Valentine, Harris Cooper (2002). Effectiveness of Mentoring Programs for Youth: A Meta-Analytic Review. *American Journal of Community Psychology* 30 (2): 157-197.

Duckworth, Angela L., Christopher Peterson, Michael D. Matthews, Dennis R. Kelly (2007). Grit: Perseverance and Passion for Long-Term Goals. *Journal of Personality and Social Psychology* 92 (6): 1087-1101.

Eby, Lillian T., Tammy D. Allen, Sarah C. Evans, Thomas Ng, David L. DuBois (2008). Does Mentoring Matter? A Multidisciplinary Meta-Analysis Comparing Mentored and Non-Mentored Individuals. *Journal of Vocational Behavior* 72 (2): 254-267.

Fersterer, Josef, Jörn-Steffen Pischke, Rudolf Winter-Ebmer (2008). Returns to Apprenticeship Training in Austria: Evidence from Failed Firms. *Scandinavian Journal of Economics* 110 (4): 733-753.

Figlio, David, Paola Giuliano, Umut Özek, Paola Sapienza (2019). Long-Term Orientation and Educational Performance. *American Economic Journal: Economic Policy* 11 (4): 272-309.

García, Jorge Luis, James J. Heckman, Duncan Ermini Leaf, María José Prados (2020). Quantifying the Life-Cycle Benefits of an Influential Early-Childhood Program. *Journal of Political Economy* 128 (7): 2502-2541.

Golsteyn, Bart H.H., Hans Grönqvist, Lena Lindahl (2014). Adolescent Time Preferences Predict Lifetime Outcomes. *Economic Journal* 124 (580): F739-F761.

Goodman, Robert (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry* 38 (5): 581-586.

Grossman, Jean Baldwin, Joseph P. Tierney (1998). Does Mentoring Work? An Impact Study of the Big Brothers Big Sisters Program. *Evaluation Review* 22 (3): 403-426.

Guryan, Jonathan, Sandra Christenson, Ashley Cureton, Ijun Lai, Jens Ludwig, Catherine Schwarz, Emma Shirey, Mary Clair Turner (2021). The Effect of Mentoring on School Attendance and Academic Outcomes: A Randomized Evaluation of the Check & Connect Program. *Journal of Policy Analysis and Management* 40 (3): 841-882.

Hanushek, Eric A, Lavinia Kinne, Philipp Lergetporer, Ludger Woessmann (2022). Patience, Risk-Taking, and Human Capital Investment across Countries. *Economic Journal* 132 (646): 2290-2307.

Hanushek, Eric A., Guido Schwerdt, Simon Wiederhold, Ludger Woessmann (2015). Returns to Skills around the World: Evidence from PIAAC. *European Economic Review* 73: 103-130.

Heckman, James J. (2008). Schools, Skills, and Synapses. *Economic Inquiry* 46 (3): 289-324.

Heckman, James J. (2020). Randomization and Social Policy Evaluation Revisited. NBER Technical Working Paper 107 (revised). Cambridge, MA: National Bureau of Economic Research.

Heckman, James J., Rodrigo Pinto (2015). Econometric Mediation Analyses: Identifying the Sources of Treatment Effects from Experimentally Estimated Production Technologies with Unmeasured and Mismeasured Inputs. *Econometric Reviews* 34 (1-2): 6-31.

Heckman, James J., Rodrigo Pinto, Peter Savelyev (2013). Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review* 103 (6): 2052-2086.

Heckman, James J., Jora Stixrud, Sergio Urzua (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics* 24 (3): 411-482.

Heller, Sara B., Anuj K. Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, Harold A. Pollack (2017). Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago. *Quarterly Journal of Economics* 132 (1): 1-54.

Herrera, Carla, Jean Baldwin Grossman, Tina J. Kauh, Jennifer McMaken (2011). Mentoring in Schools: An Impact Study of Big Brothers Big Sisters School-Based Mentoring. *Child Development* 82 (1): 346-361.

IAB (2017). Qualifikationsspezifische Arbeitslosenquoten. Nürnberg: Institut für Arbeits- und Berufsforschung.

Imbens, Guido W., Donald B. Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. New York, NY: Cambridge University Press.

Kling, Jeffrey R., Jeffrey B. Liebman, Lawrence F. Katz (2007). Experimental Analysis of Neighborhood Effects. *Econometrica* 75 (1): 83-119.

Kosse, Fabian, Thomas Deckers, Pia Pinger, Hannah Schildberg-Hörisch, Armin Falk (2020). The Formation of Prosociality: Causal Evidence on the Role of Social Environment. *Journal of Political Economy* 128 (2): 434-467.

Kosse, Fabian, Michela M. Tincani (2020). Prosociality Predicts Labor Market Success around the World. *Nature Communications* 11 (1): 5298.

Lavecchia, Adam M., Philip Oreopoulos, Robert S. Brown (2020). Long-Run Effects from Comprehensive Student Support: Evidence from Pathways to Education. *American Economic Review: Insights* 2 (2): 209-24.

Lindqvist, Erik, Roine Vestman (2011). The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment. *American Economic Journal: Applied Economics* 3 (1): 101-128.

List, John A., Azeem M. Shaikh, Yang Xu (2019). Multiple Hypothesis Testing in Experimental Economics. *Experimental Economics* 22 (4): 773-793.

Morgan, Kari Lock, Donald B. Rubin (2012). Rerandomization to Improve Covariate Balance in Experiments. *Annals of Statistics* 40 (2): 1263-1282.

OECD (2018). *A Broken Social Elevator? How to Promote Social Mobility*. Paris: OECD Publishing.

Oreopoulos, Philip, Robert S. Brown, Adam M. Lavecchia (2017). Pathways to Education: An Integrated Approach to Helping at-Risk High School Students. *Journal of Political Economy* 125 (4): 947-984.

Piopiunik, Marc, Franziska Kugler, Ludger Woessmann (2017). Einkommenserträge von Bildungsabschlüssen im Lebensverlauf: Aktuelle Berechnungen für Deutschland. *ifo Schnelldienst* 70 (7): 19-30.

Resnjanskij, Sven, Jens Ruhose, Simon Wiederhold, Ludger Woessmann (2021). Can Mentoring Alleviate Family Disadvantage in Adolescence? A Field Experiment to Improve Labor-Market Prospects. CESifo Working Paper 8870. Munich: CESifo.

Rhodes, Jean E. (2008). Improving Youth Mentoring Interventions through Research-Based Practice. *American Journal of Community Psychology* 41 (1-2): 35-42.

Rock Your Life! (2020). *Jahresbericht 2019*. München: Rock Your Life! gGmbH.

Rodríguez-Planas, Núria (2012). 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): 121-139.

Rodríguez-Planas, Núria (2014). Do Youth Mentoring Programs Change the Perspectives and Improve the Life Opportunities of at-Risk Youth? *IZA World of Labor* 2014: 62.

Romano, Joseph P., Michael Wolf (2010). Balanced Control of Generalized Error Rates. *Annals of Statistics* 38 (1): 598-633.

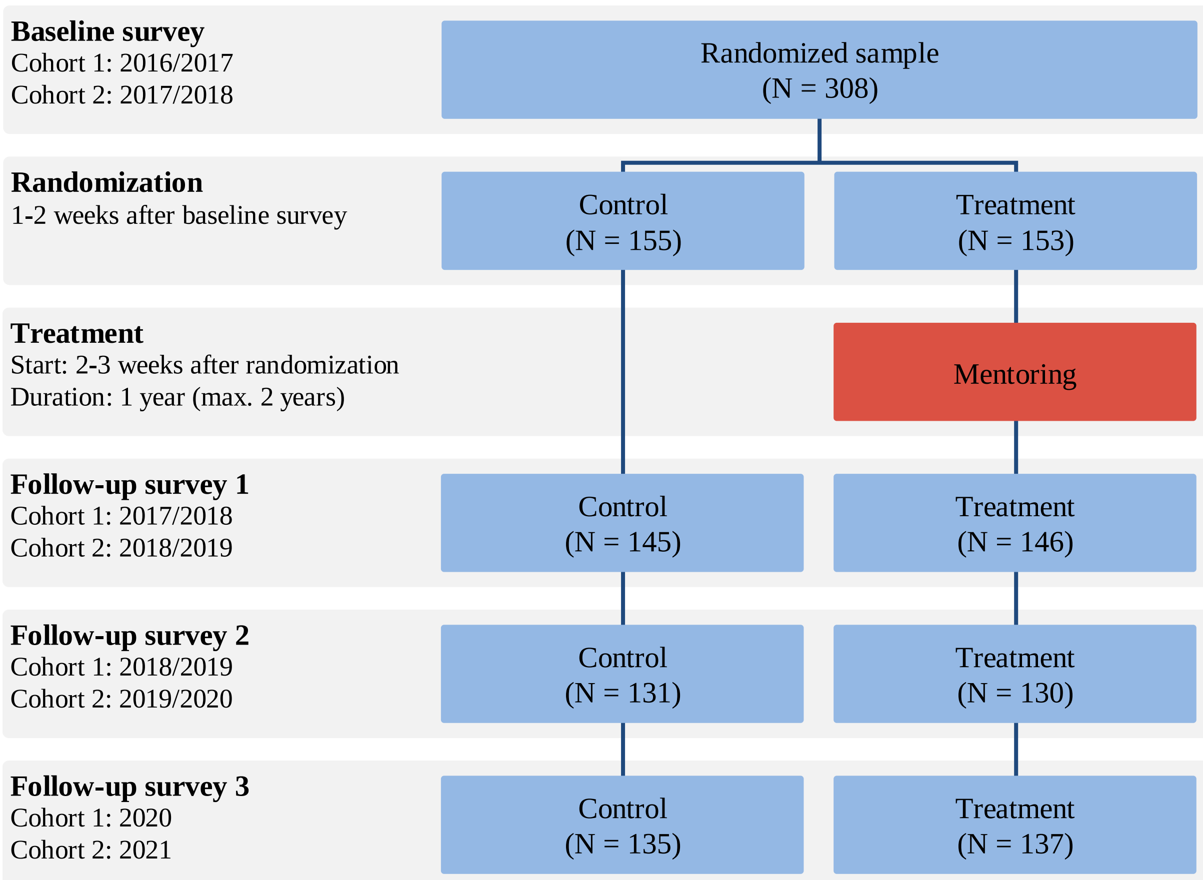
Schuetz, Gabriela, Heinrich W. Ursprung, Ludger Woessmann (2008). Education Policy and Equality of Opportunity. *Kyklos* 61 (2): 279-308.

Sutter, Matthias, Michael Weyland, Anna Untertrifaller, Manuel Froitzheim (2020). Financial Literacy, Risk and Time Preferences - Results from a Randomized Educational Intervention. CESifo Working Paper 8489. Munich: CESifo.

Westfall, Peter H., S. Stanley Young (1993). *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. New York, NY: Wiley.

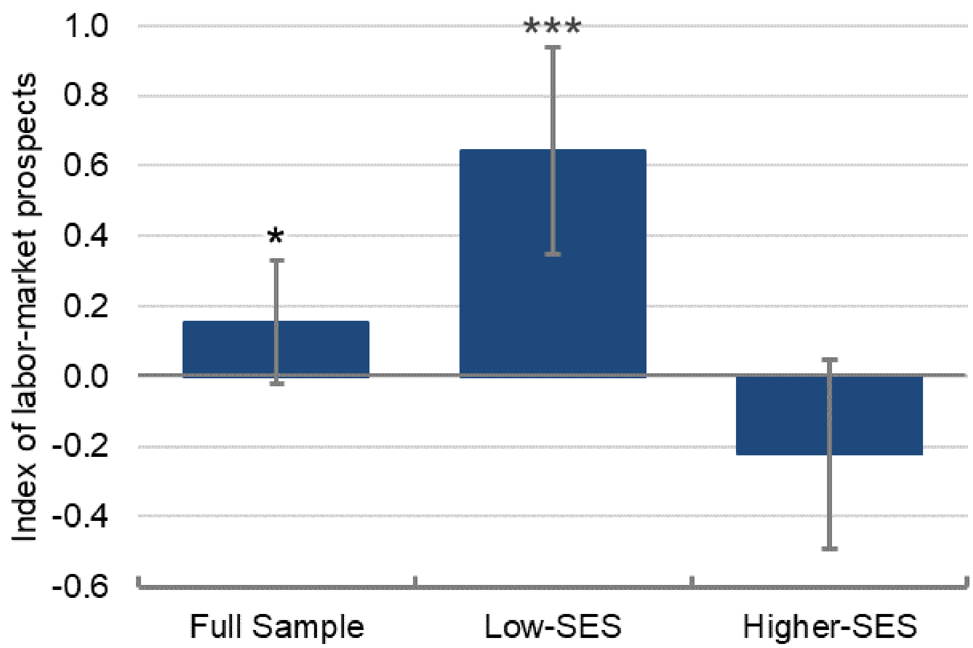
# Figures and Tables

### Figure 1: Timeline of the Surveys




Notes: Figure shows data collection and sample sizes of the randomized sample of the evaluation. Indicated sampling periods (which differ by mentoring site and cohort) refer to main field periods (sampling of over 90 percent of observations). Cohort 1 additionally included a first pilot study surveyed in the November of the year before the indicated sampling periods. See Appendix B for further information on sampling and procedural details.

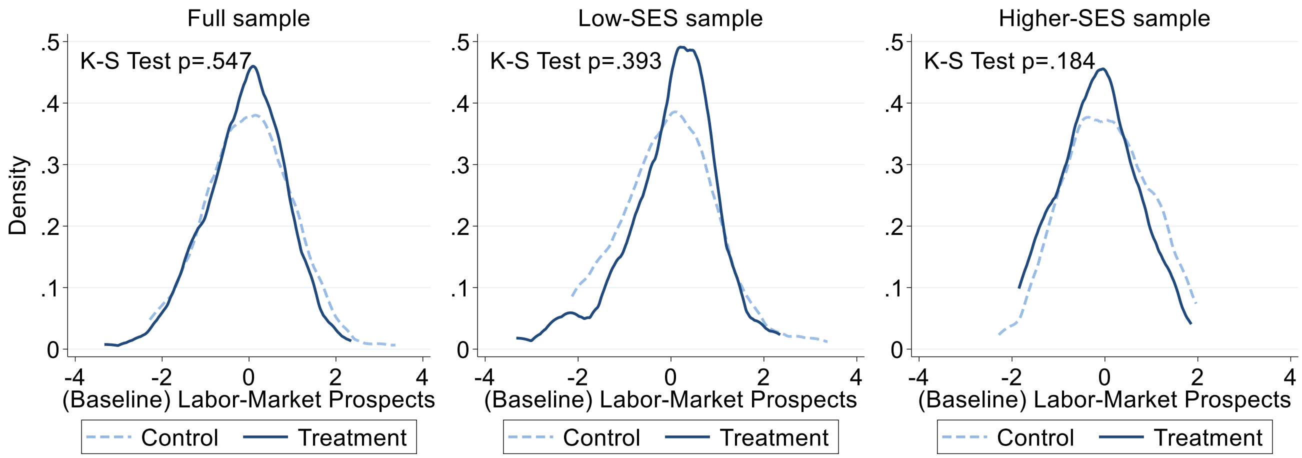
**Figure 2: Effect of the Mentoring Program on Labor-Market Prospects**



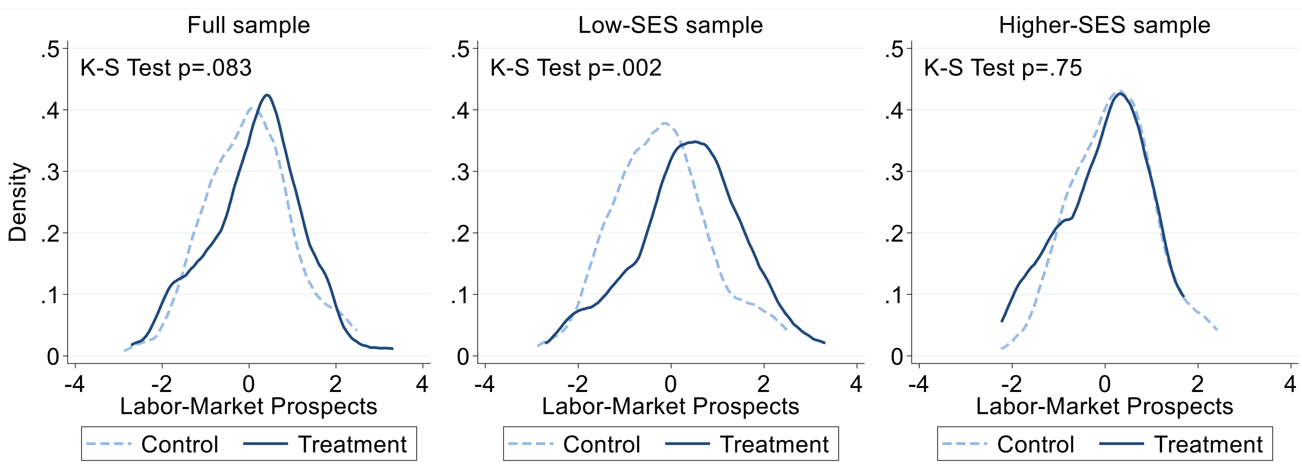

Notes: Figure shows the intention-to-treat effects (ITT) of the mentoring program on the index of labor-market prospects one year after program start, separately for all respondents (left panel), low-SES respondents (middle panel), and higher-SES respondents (right panel). See specifications in column 4 of Table 2 for details. The index of labor-market prospects is an equally weighted average of $z$-scores of three components: administrative math grade (reversed), patience and social skills index, and labor-market orientation index. Calculation of each $z$-score subtracts the score's control-group mean and divides by the control-group standard deviation. Error bars show 95 percent confidence intervals. Significance levels: $^{***}$ $p < 0.01$, $^{**}$ $p < 0.05$, $^{*}$ $p < 0.1$.

**Figure 3: Effect of the Mentoring Program on the Distribution of Labor-Market Prospects**

Panel A: Distribution of labor-market prospects at baseline



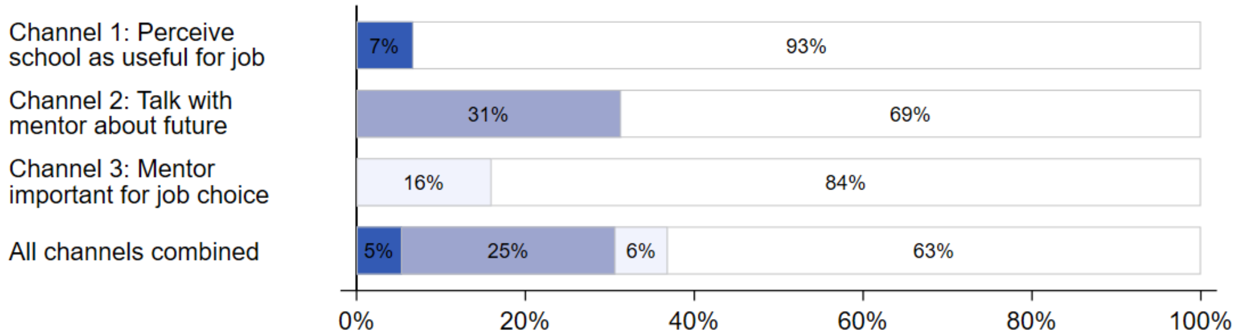

Panel B: Distribution of labor-market prospects one year after program start



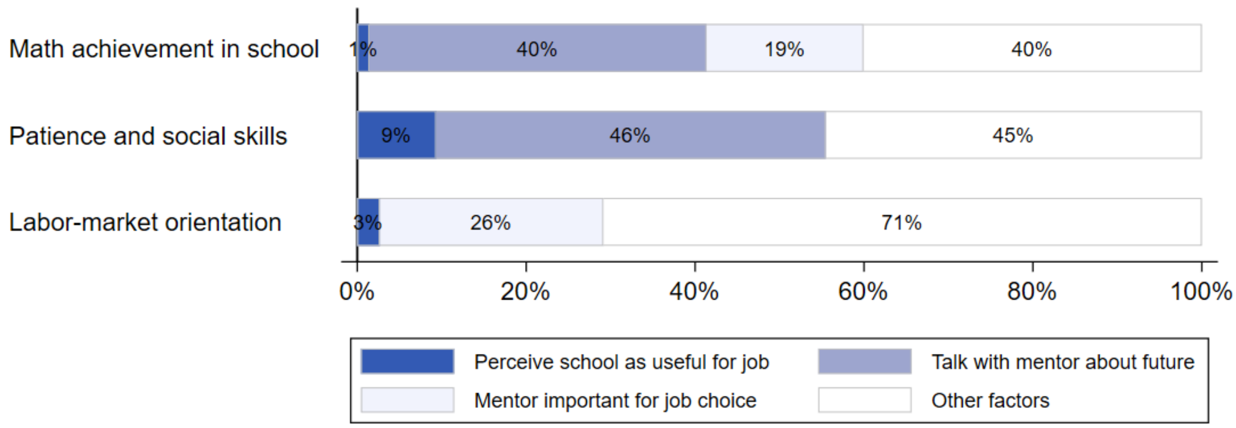

Notes: Panel A shows the entire distribution of the index of labor-market prospects for the treatment and control groups at baseline (pre-treatment). Panel B shows the unconditional treatment effect on the entire distribution of the index of labor-market prospects one year after program start. Samples: all respondents (left), low-SES respondents (middle), higher-SES respondents (right). The probability density functions are computed with an Epanechnikov kernel with bandwidth $h$ derived from the Silverman rule (Silverman (1986), pp. 47-48) with $h = 0.9An^{-1/5}$, where $n$ is the number of observations and $A = min(standard\ deviation, interquartile\ range/1.349)$. K-S Test: $p$-values for a Kolmogorov-Smirnov test of the equality of distributions of labor-market prospects of treatment and control groups.

**Figure 4: Share of Treatment Effect for Low-SES Adolescents Attributed to Mediators**

Panel A: Index of labor-market prospects

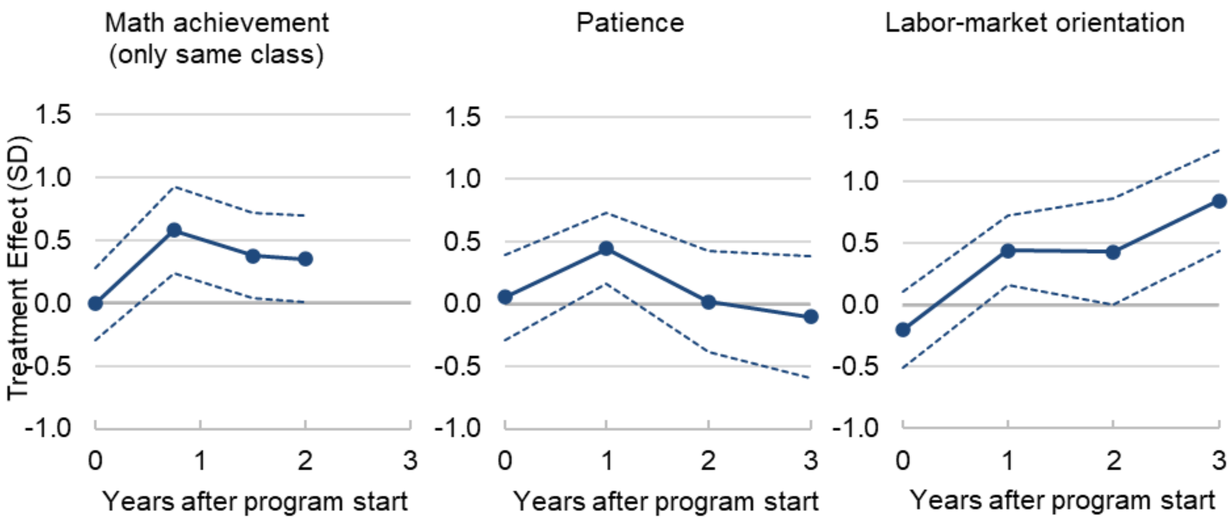



Panel B: Components




Notes: Figure shows the share of the ITT effects on the index of labor-market prospects (panel A) and on its three components (panel B) in the low-SES sample one year after program start attributed to the respective mediator in a mediation analysis. Panel B includes all channels combined (mediators with insignificant negative contributions excluded). See section 7 for details.

**Figure 5: Effect of the Mentoring Program on Components of Labor-Market Prospects for Low-SES Adolescents over Time**




Notes: Figure shows ITT effects of the mentoring program on math achievement, patience, and labor-market orientation in the low-SES sample over time. Dotted lines show 90-percent confidence bands. Covariates are from the baseline survey and include: outcome from the baseline survey, gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in $t_0$ are included. Additional controls include fixed effects for surveys conducted during first (March-May 2020) and second (Oct 2020-Feb 2021) Covid-19 wave.

**Table 1: Balancing**

| | Control | Treatment | Difference | Difference by SES | Observa-tions |
|---|---|---|---|---|---|
| | Mean | Mean | *p*-value | *p*-value | |
| | (1) | (2) | (3) | (4) | (5) |
| **A. Outcome variables at baseline** | | | | | |
| Overall index | 0.00 | -0.09 | 0.433 | 0.356 | 308 |
| *Components* | | | | | |
| Math grade (administrative) | 0.00 | 0.02 | 0.889 | 0.242 | 218 |
| Math grade (admin.) missing dummy | 0.28 | 0.30 | 0.747 | 0.912 | 308 |
| Patience and social skills index | 0.00 | -0.07 | 0.548 | 0.671 | 308 |
| Labor-market orientation index | 0.00 | -0.09 | 0.424 | 0.755 | 307 |
| **B. Components of outcome variables at baseline** | | | | | |
| *Patience and social skills index* | | | | | |
| Patience | 0.00 | -0.02 | 0.891 | 0.990 | 308 |
| Social skills index | 0.00 | -0.09 | 0.402 | 0.380 | 308 |
| Prosociality | 0.00 | 0.01 | 0.897 | 0.628 | 308 |
| Trust | 0.00 | -0.05 | 0.665 | 0.844 | 307 |
| Self-efficacy | 0.00 | -0.15 | 0.158 | 0.135 | 308 |
| *Labor-market orientation index* | | | | | |
| Wants apprenticeship after school | 0.36 | 0.37 | 0.889 | 0.902 | 267 |
| Knows future career | 0.00 | -0.16 | 0.156 | 0.393 | 307 |
| **C. Matching and balancing variables for randomization at baseline** | | | | | |
| Male | 0.43 | 0.44 | 0.921 | 0.434 | 308 |
| Age | 13.99 | 13.97 | 0.851 | 0.158 | 308 |
| Migrant | 0.59 | 0.57 | 0.744 | 0.998 | 308 |
| Books at home | 1.73 | 1.67 | 0.461 | 0.132 | 308 |
| Math grade (survey) | 1.71 | 1.73 | 0.806 | 0.639 | 261 |
| Math grade (survey) missing dummy | 0.14 | 0.16 | 0.602 | 0.934 | 308 |
| German grade (survey) | 1.73 | 1.71 | 0.751 | 0.868 | 258 |
| German grade (survey) missing dummy | 0.15 | 0.17 | 0.721 | 0.971 | 308 |
| English grade (survey) | 1.79 | 1.83 | 0.626 | 0.282 | 258 |
| English grade (survey) missing dummy | 0.15 | 0.17 | 0.721 | 0.965 | 308 |
| Received paid private teaching | 0.18 | 0.21 | 0.529 | 0.792 | 308 |
| Parental homework support | 2.81 | 2.71 | 0.368 | 0.521 | 307 |
| Big-5: Conscientiousness | 3.35 | 3.26 | 0.327 | 0.268 | 308 |
| Big-5: Neuroticism | 2.91 | 2.98 | 0.413 | 0.466 | 308 |
| **D. Further control variables at baseline** | | | | | |
| Big-5: Openness | 3.41 | 3.51 | 0.337 | 0.199 | 308 |
| Big-5: Extraversion | 3.31 | 3.35 | 0.610 | 0.645 | 308 |
| Big-5: Agreeableness | 3.50 | 3.46 | 0.704 | 0.765 | 307 |
| Higher-SES | 0.53 | 0.53 | 0.357 | – | 308 |

Notes: Table shows group means after randomization for control group (column 1) and treatment group (column 2) in the baseline survey. Sample consists of all respondents in the matched pairs. Column 3 shows the *p*-value of the coefficient on the treatment indicator in a regression of the specific variable on the treatment indicator. Column 4 shows the *p*-value of an *F*-test of joint significance of the coefficients on the treatment indicator and the treatment indicator interacted with the higher-SES dummy in a regression of the specific variable on the treatment indicator, the higher-SES dummy, and their interaction.

**Table 2: Effect of the Mentoring Program on Index of Labor-Market Prospects**

| | ITT | | | | TOT |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| **Panel A: Average treatment effects** | | | | | |
| Treatment | 0.149 | 0.199$^{**}$ | 0.193$^{**}$ | 0.153$^{*}$ | 0.177$^{*}$ |
| | *(0.119)* | *(0.100)* | *(0.091)* | *(0.089)* | *(0.103)* |
| | *[0.140]* | *[0.030]* | *[0.035]* | *[0.118]* | – |
| Outcome in $t_0$ | | 0.585$^{***}$ | 0.480$^{***}$ | 0.441$^{***}$ | 0.429$^{***}$ |
| | | *(0.051)* | *(0.065)* | *(0.082)* | *(0.083)* |
| $R^2$ | 0.005 | 0.295 | 0.703 | 0.730 | 0.731 |
| Kleibergen-Paap $F$ statistic | | | | | 858.65 |
| **Panel B: Heterogeneous treatment effects by SES** | | | | | |
| Treatment | 0.563$^{***}$ | 0.544$^{***}$ | 0.618$^{***}$ | 0.644$^{***}$ | 0.771$^{***}$ |
| | *(0.194)* | *(0.151)* | *(0.140)* | *(0.149)* | *(0.175)* |
| | *[0.001]* | *[0.000]* | *[0.000]* | *[0.000]* | – |
| Treatment x Higher-SES | -0.757$^{***}$ | -0.632$^{***}$ | -0.755$^{***}$ | -0.865$^{***}$ | -1.016$^{***}$ |
| | *(0.242)* | *(0.199)* | *(0.206)* | *(0.227)* | *(0.260)* |
| | *[0.001]* | *[0.001]* | *[0.000]* | *[0.000]* | – |
| Higher-SES | 0.361$^{**}$ | 0.313$^{**}$ | 0.194 | 0.286 | 0.288 |
| | *(0.161)* | *(0.143)* | *(0.186)* | *(0.192)* | *(0.195)* |
| Outcome in $t_0$ | | 0.576$^{***}$ | 0.495$^{***}$ | 0.446$^{***}$ | 0.429$^{***}$ |
| | | *(0.052)* | *(0.065)* | *(0.076)* | *(0.078)* |
| $R^2$ | 0.038 | 0.318 | 0.728 | 0.758 | 0.756 |
| Kleibergen-Paap $F$ statistic | | | | | 197.22 |
| Treatment effect for Higher-SES | -0.194 | -0.088 | -0.137 | -0.221 | -0.245 |
| | *(0.144)* | *(0.131)* | *(0.132)* | *(0.136)* | *(0.155)* |
| Randomization-pair fixed effects | No | No | Yes | Yes | Yes |
| Covariates | No | No | No | Yes | Yes |
| Observations | 304 | 304 | 304 | 304 | 304 |

Notes: Table shows intention-to-treat (ITT) effects and treatment-on-treated (TOT) effects of the mentoring program on the index of labor-market prospects one year after program start. Panel A shows average treatment effects and panel B shows treatment effects by socioeconomic status. The index of labor-market prospects is an equally weighted average of $z$-scores of three components: administrative math grade (reversed), patience and social skills index, and labor-market orientation index. Calculation of each $z$-score subtracts the score's control-group mean and divides by the control-group standard deviation. Columns 1-4: ordinary least squares estimates; column 5: two-stage least squares estimates. In the TOT estimation in column 5, *Treatment* indicates program take-up (one if mentor and mentee have met at least once, zero otherwise), which is instrumented by the random treatment assignment. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in $t_0$ are included. Robust standard errors in parentheses. Randomization inference (RI) $p$-values in square brackets, obtained from RI with 1,000 permutations, assigning the treatment status randomly within randomization pairs. Significance levels: $^{***}$ $p < 0.01$, $^{**}$ $p < 0.05$, $^{*}$ $p < 0.1$.

**Table 3: Effect of the Mentoring Program on Components of the Index of Labor-Market Prospects**

| | Math grade | | Patience and social skills | | | Labor-market orientation |
|---|---|---|---|---|---|---|
| | All | Same class only | Index | Patience | Social skills | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment | 0.309** | 0.463*** | 0.468*** | 0.444** | 0.280 | 0.443*** |
| | *(0.144)* | *(0.169)* | *(0.166)* | *(0.172)* | *(0.192)* | *(0.169)* |
| | *[0.042]* | *[0.003]* | [0.001] | [0.010] | [0.138] | [0.009] |
| Treatment x Higher-SES | -0.493** | -0.556** | -0.593** | -0.506* | -0.403 | -0.544** |
| | *(0.218)* | *(0.256)* | *(0.260)* | *(0.278)* | *(0.251)* | *(0.273)* |
| | *[0.026]* | *[0.042]* | [0.012] | [0.052] | [0.098] | [0.034] |
| Higher-SES | 0.305 | 0.152 | 0.072 | 0.103 | 0.008 | 0.192 |
| | *(0.196)* | *(0.218)* | *(0.227)* | *(0.228)* | *(0.223)* | *(0.227)* |
| Outcome in $t_0$ | 0.497*** | 0.600*** | 0.268*** | 0.254*** | 0.474*** | 0.376*** |
| | *(0.099)* | *(0.093)* | (0.102) | (0.089) | (0.117) | (0.089) |
| Randomization-pair fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 294 | 221 | 291 | 291 | 291 | 291 |
| $R^2$ | 0.777 | 0.768 | 0.698 | 0.648 | 0.706 | 0.699 |
| Treatment effect for Higher-SES | -0.184 | -0.093 | -0.126 | -0.062 | -0.122 | -0.101 |
| | *(0.134)* | *(0.166)* | *(0.165)* | *(0.180)* | *(0.133)* | *(0.170)* |
| SES gap | 0.436 | 0.393 | 0.237 | 0.131 | 0.228 | -0.018 |

Notes: Table shows ITT effects of the mentoring program on administrative math grades (columns 1 and 2), patience and social skills (columns 3-5), and labor-market orientation (column 6) one year after program start. Columns 1 and 2: grades are standardized by subtracting the control-group mean and dividing by the control-group standard deviation; order of grades is reversed so that higher values indicate better outcomes. Columns 3-6: variables and indices are standardized by subtracting the control-group mean and dividing by the control-group standard deviation. Ordinary least squares estimates. *SES gap* is calculated as the coefficient on higher-SES background in a regression of the respective outcome on the higher-SES indicator in the control-group sample one year after program start. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in $t_0$ are included. Robust standard errors in parentheses. Randomization inference (RI) *p*-values in square brackets, obtained from RI with 1,000 permutations, assigning the treatment status randomly within randomization pairs. Significance levels: $^{***}$ $p < 0.01$, $^{**}$ $p < 0.05$, $^{*}$ $p < 0.1$.

**Table 4: Effect of the Mentoring Program on Index of Labor-Market Prospects for each of the SES Components**

| | Multidimensional SES measure | Components of SES measure | | | |
|---|---|---|---|---|---|
| | | Books at home | Parental education | Single-parent status | First-generation migrant |
| | (1) | (2) | (3) | (4) | (5) |
| Treatment | 0.644*** | 0.556*** | 0.260** | 0.240 | 0.753*** |
| | *(0.149)* | *(0.143)* | *(0.118)* | *(0.228)* | *(0.281)* |
| | *[0.000]* | *[0.000]* | *[0.036]* | *[0.332]* | *[0.008]* |
| Treatment x Higher-SES | -0.865*** | -0.748*** | -0.317 | -0.127 | -0.697** |
| | *(0.227)* | *(0.220)* | *(0.269)* | *(0.284)* | *(0.321)* |
| | *[0.000]* | *[0.002]* | *[0.234]* | *[0.654]* | *[0.036]* |
| Higher-SES | 0.286 | 0.182 | 0.246 | -0.034 | 0.466* |
| | *(0.192)* | *(0.195)* | *(0.212)* | *(0.228)* | *(0.261)* |
| Outcome in $t_0$ | 0.446*** | 0.459*** | 0.456*** | 0.442*** | 0.434*** |
| | *(0.076)* | *(0.078)* | *(0.084)* | *(0.083)* | *(0.080)* |
| Randomization-pair fixed effects | Yes | Yes | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes | Yes | Yes |
| Observations | 304 | 304 | 304 | 304 | 304 |
| $R^2$ | 0.758 | 0.753 | 0.734 | 0.732 | 0.738 |
| Treatment effect for Higher-SES | -0.221 | -0.192 | -0.056 | 0.113 | 0.056 |
| | *(0.136)* | *(0.137)* | *(0.210)* | *(0.116)* | *(0.102)* |
| SES gap | 0.361 | 0.476 | 0.128 | 0.345 | -0.008 |
| Share with low-SES status | 0.46 | 0.47 | 0.67 | 0.25 | 0.13 |

Notes: Table shows ITT effects of the mentoring program on the index of labor-market prospects one year after program start. The index is an equally weighted average of *z*-scores of three components: administrative math grade (reversed), patience and social skills index, and labor-market orientation index. Calculation of each *z*-score subtracts the score's control-group mean and divides by the control-group standard deviation. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in $t_0$ are included. Robust standard errors in parentheses. Randomization inference (RI) *p*-values in square brackets, obtained from RI with 1,000 permutations, assigning the treatment status randomly within randomization pairs. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

**Table 5: Effect of the Mentoring Program on Satisfaction Outcomes**

| | Life | Own money and wealth | School | Performance in math |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Treatment | 0.246** | 0.137* | 0.155 | 0.102 |
| | *(0.102)* | *(0.079)* | *(0.101)* | *(0.106)* |
| | *[0.023]* | *[0.077]* | *[0.117]* | *[0.327]* |
| Treatment x Higher-SES | -0.295** | -0.176 | -0.227 | -0.075 |
| | *(0.141)* | *(0.109)* | *(0.153)* | *(0.150)* |
| | *[0.043]* | *[0.078]* | *[0.119]* | *[0.596]* |
| Higher-SES | 0.095 | 0.121 | 0.004 | 0.031 |
| | *(0.105)* | *(0.086)* | *(0.120)* | *(0.121)* |
| Outcome in $t_0$ | 0.163 | 0.199** | 0.348*** | 0.232*** |
| | *(0.101)* | *(0.090)* | *(0.098)* | *(0.084)* |
| Randomization-pair fixed effects | Yes | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes | Yes |
| Observations | 291 | 291 | 291 | 291 |
| $R^2$ | 0.637 | 0.635 | 0.597 | 0.632 |
| Treatment effect for Higher-SES | -0.049 | -0.039 | -0.072 | 0.027 |
| | *(0.073)* | *(0.055)* | *(0.094)* | *(0.084)* |

Notes: Table shows ITT effects of the mentoring program on satisfaction domains (elicited one year after program start) indicated in the column header. Satisfaction in each domain is reported on a 5-point Likert scale, ranging from “totally dissatisfied” to “totally satisfied”. Dependent variables are dummies, which are one if individuals report that they are “somewhat satisfied” or “totally satisfied”, and zero otherwise. Column 1: How satisfied are you currently, all in all, with your life? Column 2: How satisfied are you with your belongings? Think about money and things that you own. Column 3: How satisfied are you with your situation in school? Column 4: How satisfied are you with your performance in math. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in $t_0$ are included. Robust standard errors in parentheses. Randomization inference (RI) *p*-values in square brackets, obtained from RI with 1,000 permutations, assigning the treatment status randomly within randomization pairs. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

**Table 6: Effect of the Mentoring Program on Early Realizations of Labor-Market Transitions**

| | Early career realizations | | | | Satisfaction with current situation | |
|---|---|---|---|---|---|---|
| | Doing an apprenticeship | Attending school | Work-related activities | Unemployed and other | Satisfied | No desire to change |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment | 0.293** | -0.205 | 0.029 | -0.117 | 0.312*** | 0.220 |
| | *(0.118)* | *(0.129)* | *(0.090)* | *(0.078)* | *(0.115)* | *(0.137)* |
| | *[0.006]* | *[0.106]* | *[0.739]* | *[0.136]* | *[0.002]* | *[0.072]* |
| Treatment x Higher-SES | -0.439** | 0.274 | -0.015 | 0.180* | -0.376** | -0.133 |
| | *(0.171)* | *(0.174)* | *(0.129)* | *(0.094)* | *(0.147)* | *(0.183)* |
| | *[0.007]* | *[0.094]* | *[0.912]* | *[0.060]* | *[0.005]* | *[0.423]* |
| Higher-SES | 0.195 | -0.060 | -0.019 | -0.115* | 0.148 | 0.089 |
| | *(0.123)* | *(0.120)* | *(0.086)* | *(0.064)* | *(0.128)* | *(0.131)* |
| Randomization-pair fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 272 | 272 | 272 | 272 | 272 | 272 |
| $R^2$ | 0.607 | 0.696 | 0.598 | 0.747 | 0.708 | 0.622 |
| Treatment effect for Higher-SES | -0.147 | 0.069 | 0.014 | 0.063 | -0.063 | 0.088 |
| | *(0.094)* | *(0.087)* | *(0.066)* | *(0.036)* | *(0.069)* | *(0.094)* |
| SES gap | 0.000 | 0.117 | -0.020 | -0.097 | 0.067 | 0.157 |
| Control-group mean | 0.27 | 0.55 | 0.09 | 0.10 | 0.70 | 0.64 |
| Low-SES control-group mean | 0.27 | 0.48 | 0.10 | 0.15 | 0.67 | 0.55 |

Notes: Table shows ITT effects of the mentoring program on early career realizations and satisfaction with the current situation, elicited three years after program start. All outcomes are dummy variables. Column 2: same school, higher-track school, or school-based pre-employment training. Column 3: working, internship, or voluntary social year. Column 5: respondent satisfied with current situation. Column 6: respondent has no desire to change current situation. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, Big-5 personality traits (all columns) as well as labor-market orientation index in columns 1-4 and satisfaction with life in columns 5 and 6. Dummies for missing values in $t_0$ are included. Additional controls include fixed effects for surveys conducted during first (March-May 2020) and second (Oct 2020-Feb 2021) Covid-19 wave. *SES gap* is calculated as the coefficient on higher-SES background in a regression of the respective outcome on the higher-SES indicator in the control-group sample in the survey three years after program start. Robust standard errors in parentheses. Randomization inference (RI) *p*-values in square brackets, obtained from RI with 1,000 permutations, assigning the treatment status randomly within randomization pairs. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

# Appendix (for online publication only)

# Can Mentoring Alleviate Family Disadvantage in Adolescence? A Field Experiment to Improve Labor-Market Prospects

Sven Resnjanskij, Jens Ruhose, Simon Wiederhold, Ludger Woessmann, and Katharina Wedel[†]

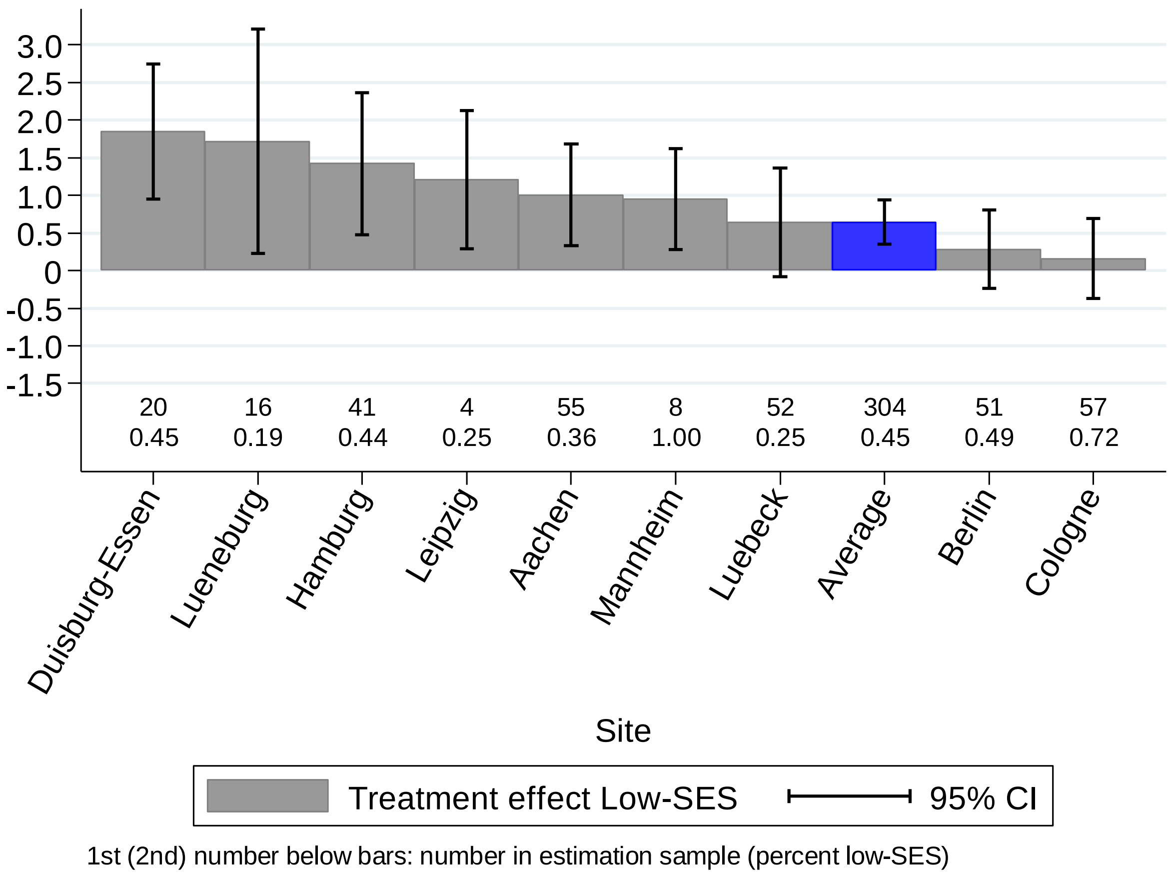
September 25, 2022



[†] Resnjanskij: ifo Institute at the University of Munich and CESifo, resnjanskij@ifo.de; Ruhose: Kiel University, CESifo, and IZA, ruhose@economics.uni-kiel.de; Wiederhold: Catholic University Eichstaett-Ingolstadt, CESifo, and ROA, simon.wiederhold@ku.de; Woessmann: University of Munich, ifo Institute, CESifo, and IZA, woessmann@ifo.de; Wedel: ifo Institute at the University of Munich, wedel@ifo.de.

## Appendix A: Additional Figures and Tables

**Figure A1: Site-specific Effects of the Mentoring Program on Labor-Market Prospects**




Notes: Figure shows the intention-to-treat effects (ITT) of the mentoring program on the index of labor-market prospects for low-SES adolescents one year after program start. Site-specific treatment effects are obtained by adding a triple interaction between treatment, SES, and indicators for each specific mentoring site. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in $t_0$ are included. Error bars show 95 percent confidence intervals obtained from robust standard errors. Significance levels of differences: $^{***}$ $p < 0.01$, $^{**}$ $p < 0.05$, $^{*}$ $p < 0.1$.

## Table A1: Comparison of Sample to German Student Population in PISA

| | All | | Low-SES | | | Higher-SES | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Δ PISA total | Mean | Δ PISA total | Δ PISA low-SES | Mean | Δ PISA total | Δ PISA higher-SES |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Low-SES | 0.46 | 0.22*** | 1.00 | 0.76*** | — | 0.00 | -0.24*** | — |
| *Books at home* | | | | | | | | |
| 0-10 books | 0.24 | 0.15*** | 0.48 | 0.39*** | 0.11** | 0.04 | -0.06** | 0.03*** |
| 11-25 books | 0.23 | 0.10*** | 0.43 | 0.29*** | -0.05 | 0.06 | -0.07** | 0.04*** |
| 26-100 books | 0.25 | -0.03 | 0.05 | -0.23*** | -0.02 | 0.43 | 0.14*** | 0.07* |
| 101-200 books | 0.11 | -0.10*** | 0.00 | -0.20*** | -0.03** | 0.20 | -0.01 | -0.06* |
| 201-500 books | 0.09 | -0.09*** | 0.03 | -0.15*** | -0.01 | 0.14 | -0.03 | -0.08** |
| More than 500 books | 0.08 | -0.03 | 0.01 | -0.09*** | 0.00 | 0.13 | 0.02 | 0.00 |
| *Student demographics* | | | | | | | | |
| Male | 0.44 | -0.06** | 0.47 | -0.03 | -0.07 | 0.41 | -0.09** | -0.08* |
| Age | 13.98 | -1.85*** | 14.28 | -1.55*** | -1.56*** | 13.72 | -2.10*** | -2.10*** |
| Migrant | 0.58 | 0.30*** | 0.74 | 0.46*** | 0.23*** | 0.44 | 0.16*** | 0.23*** |
| First-generation migrant | 0.13 | 0.09*** | 0.28 | 0.24*** | 0.11*** | 0.00 | -0.04** | — |
| Single-parenthood status | 0.25 | 0.12*** | 0.32 | 0.18*** | 0.14*** | 0.20 | 0.06** | 0.07*** |
| *University degree, parents* | | | | | | | | |
| Yes, at least one parent | 0.31 | 0.01 | 0.15 | -0.15*** | 0.06*** | 0.46 | 0.15*** | 0.09** |
| Missing (for both parents) | 0.26 | 0.19*** | 0.29 | 0.22*** | 0.13*** | 0.23 | 0.16*** | 0.19*** |
| *University degree, father* | | | | | | | | |
| Yes | 0.22 | -0.01 | 0.12 | -0.11*** | 0.06** | 0.31 | 0.07** | 0.02 |
| Missing | 0.40 | 0.26*** | 0.45 | 0.32*** | 0.19*** | 0.35 | 0.22*** | 0.26*** |
| *University degree, mother* | | | | | | | | |
| Yes | 0.23 | 0.06*** | 0.09 | -0.09** | 0.02 | 0.35 | 0.18*** | 0.14*** |
| Missing | 0.32 | 0.22*** | 0.33 | 0.23*** | 0.11*** | 0.32 | 0.22*** | 0.25*** |
| *Employment, father* | | | | | | | | |
| Full-time | 0.70 | -0.11*** | 0.67 | -0.13*** | -0.06 | 0.72 | -0.08** | -0.11*** |
| Part-time | 0.06 | -0.01 | 0.02 | -0.05** | -0.05** | 0.09 | 0.02 | 0.02 |
| Not employed, not searching | 0.03 | -0.01 | 0.04 | 0.00 | 0.00 | 0.02 | -0.02 | -0.02 |
| Unemployed | 0.03 | 0.00 | 0.04 | 0.02 | 0.00 | 0.01 | -0.01 | -0.01 |
| Missing | 0.19 | 0.12*** | 0.22 | 0.16*** | 0.11*** | 0.16 | 0.09*** | 0.11*** |
| *Employment, mother* | | | | | | | | |
| Full-time | 0.39 | 0.07*** | 0.35 | 0.03 | 0.04 | 0.43 | 0.11*** | 0.10*** |
| Part-time | 0.25 | -0.17*** | 0.25 | -0.17*** | -0.09** | 0.25 | -0.17*** | -0.19*** |
| Not employed, not searching | 0.21 | 0.03 | 0.25 | 0.07* | 0.01 | 0.18 | 0.00 | 0.01 |
| Unemployed | 0.06 | 0.02* | 0.05 | 0.01 | 0.00 | 0.07 | 0.03** | 0.03** |
| Missing | 0.09 | 0.05*** | 0.11 | 0.07*** | 0.04 | 0.07 | 0.03** | 0.04*** |

Notes: Table shows group means for our baseline sample ($N$=308) and differences to the PISA 2012 sample. Columns 1-2: full sample; columns 3-5: low-SES sample; columns 6-8: higher-SES sample. Significance of the difference is tested by a two-side $t$-test on the mean. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

## Table A2: Balancing in Subsamples

| | Low-SES sample | | | Higher-SES sample | | | (1) vs. (4) |
|---|---|---|---|---|---|---|---|
| | Control Mean | Treatment Mean | Difference *p*-value | Control Mean | Treatment Mean | Difference *p*-value | Difference *p*-value |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| **A. Outcome variables at baseline** | | | | | | | |
| Overall index | -0.05 | 0.00 | 0.795 | 0.05 | -0.15 | 0.158 | 0.566 |
| *Components* | | | | | | | |
| Math grade (administrative) | -0.16 | 0.08 | 0.218 | 0.18 | -0.03 | 0.254 | 0.073 |
| Math grade (admin.) missing d. | 0.23 | 0.26 | 0.672 | 0.34 | 0.33 | 0.955 | 0.126 |
| Patience and social skills index | -0.09 | -0.09 | 0.982 | 0.08 | -0.05 | 0.373 | 0.288 |
| Labor-market orientation index | 0.14 | -0.03 | 0.504 | -0.13 | -0.18 | 0.737 | 0.101 |
| **B. Components of outcome variables at baseline** | | | | | | | |
| *Patience and social skills index* | | | | | | | |
| Patience | 0.01 | -0.02 | 0.898 | -0.01 | -0.02 | 0.953 | 0.925 |
| Social skills index | -0.14 | -0.13 | 0.913 | 0.13 | -0.06 | 0.167 | 0.086 |
| *Components* | | | | | | | |
| Prosociality | -0.08 | 0.05 | 0.444 | 0.07 | -0.01 | 0.558 | 0.346 |
| Trust | -0.09 | -0.19 | 0.587 | 0.09 | 0.06 | 0.837 | 0.267 |
| Self-efficacy | -0.13 | -0.13 | 0.995 | 0.12 | -0.18 | 0.046 | 0.130 |
| *Labor-market orientation index* | | | | | | | |
| Wants apprenticeship after school | 0.45 | 0.44 | 0.934 | 0.29 | 0.32 | 0.656 | 0.054 |
| Knows future career | 0.07 | -0.09 | 0.332 | -0.06 | -0.21 | 0.338 | 0.412 |
| **C. Matching and balancing variables for randomization at baseline** | | | | | | | |
| Male | 0.43 | 0.52 | 0.297 | 0.44 | 0.38 | 0.448 | 0.893 |
| Age | 14.17 | 14.39 | 0.183 | 13.81 | 13.64 | 0.168 | 0.013 |
| Migrant | 0.75 | 0.74 | 0.954 | 0.44 | 0.44 | 0.993 | 0.000 |
| Books at home | 1.19 | 1.08 | 0.133 | 2.24 | 2.12 | 0.182 | 0.000 |
| Math grade (survey) | 1.80 | 1.75 | 0.696 | 1.63 | 1.72 | 0.390 | 0.125 |
| Math grade (survey) missing d. | 0.08 | 0.09 | 0.819 | 0.20 | 0.22 | 0.772 | 0.031 |
| German grade (survey) | 1.77 | 1.78 | 0.895 | 1.69 | 1.64 | 0.607 | 0.507 |
| German grade (survey) missing d. | 0.08 | 0.09 | 0.819 | 0.23 | 0.23 | 0.940 | 0.011 |
| English grade (survey) | 1.88 | 2.07 | 0.146 | 1.70 | 1.63 | 0.526 | 0.113 |
| English grade (survey) missing d. | 0.09 | 0.11 | 0.803 | 0.21 | 0.22 | 0.927 | 0.039 |
| Received paid private teaching | 0.17 | 0.18 | 0.896 | 0.19 | 0.23 | 0.503 | 0.820 |
| Parental homework support | 2.57 | 2.46 | 0.501 | 3.03 | 2.90 | 0.357 | 0.003 |
| Big-5: Conscientiousness | 3.47 | 3.24 | 0.111 | 3.25 | 3.28 | 0.796 | 0.065 |
| Big-5: Neuroticism | 2.89 | 2.86 | 0.814 | 2.93 | 3.07 | 0.226 | 0.774 |
| **D. Further control variables at baseline** | | | | | | | |
| Big-5: Openness | 3.38 | 3.30 | 0.612 | 3.44 | 3.67 | 0.085 | 0.695 |
| Big-5: Extraversion | 3.26 | 3.39 | 0.363 | 3.35 | 3.32 | 0.832 | 0.517 |
| Big-5: Agreeableness | 3.51 | 3.55 | 0.762 | 3.49 | 3.40 | 0.505 | 0.925 |
| Higher-SES | 0.00 | 0.00 | | 1.00 | 1.00 | | |

Notes: Table shows group means after randomization for control and treatment group by SES sample in the baseline survey. Sample consists of all adolescents in the matched pairs. Columns 3 and 6 show the *p*-value of the coefficient on the treatment indicator in a regression that regresses the specific variable on the treatment indicator in the respective sample. Column 7 shows the *p*-value of the coefficient on the higher-SES indicator in a regression that regresses the specific variable on the higher-SES indicator in the control group.

## Table A3: Variable Definitions and Wording of Questionnaire Items

| | Wording (English translation) (1) | Wording (German original) (2) | Answer categories (3) |
|---|---|---|---|
| **Outcome variables (follow-up surveys)** | | | |
| *Index of labor-market prospects* | *Equally weighted average of three components: administrative math grade (reversed); patience and social skills index; labor-market orientation index* | | |
| Math grade (administrative) | Administrative math grade in school (standardized by subtracting control-group mean and dividing by control-group standard deviation separately by survey round) | – | Ordering reversed so that higher values indicate better outcome |
| *Patience and social skills index* | *Equally weighted average of two components: patience; social skills index* | | |
| Patience | Agreement to three items (German SOEP): I abstain from things today to be able to afford more tomorrow; I prefer to have fun today and don't think about tomorrow (reversed); I tend to postpone things until later, even if it would be better to do them immediately (reversed). | Ich verzichte heute auf etwas, damit ich mir morgen mehr leisten kann; Ich will heute meinen Spaß haben und denke dabei nicht an morgen; Ich neige dazu, Dinge auf später zu verschieben, auch wenn es besser wäre, diese sofort zu erledigen. | 5-point scales from "does not apply at all" to "applies completely" |
| *Social skills index* | *Equally weighted average of three components: prosociality; trust; self-efficacy* | | |
| Prosociality | Agreement to five items (Strength and Difficulties Questionnaire, SDQ, Goodman (1997)): I try to be nice to other people, I care about their feelings; I usually share with others (sweets, toys, crayons, etc.); I am helpful if someone is hurt, ill or upset; I am kind to younger children; I often volunteer to help others (parents, teachers, children). | Ich versuche, nett zu anderen Menschen zu sein, ihre Gefühle sind mir wichtig; Ich teile normalerweise mit anderen (Süßigkeiten, Spielzeug, Buntstifte usw.); Ich bin hilfsbereit, wenn andere verletzt, krank oder betrübt sind; Ich bin nett zu jüngeren Kindern; Ich helfe anderen oft freiwillig (Eltern, Lehrern oder Gleichaltrigen). | 3-point scales: does not apply; applies partially; applies completely |
| Trust | Agreement to item: In general, one can trust people. | Im Allgemeinen kann man den Menschen vertrauen. | 11-point scale from "does not apply at all" to "applies completely" |
| Self-efficacy | Agreement to three items (General Self-efficacy Short Scale, Beierlein et al. (2012)): In difficult situations, I can trust in my abilities; I am able to solve most problems on my own; I can usually solve even challenging and complex tasks well. | Allgemeine Selbstwirksamkeit Kurzskala: In schwierigen Situationen kann ich mich auf meine Fähigkeiten verlassen; Die meisten Probleme kann ich aus eigener Kraft gut meistern; Auch anstrengende und komplizierte Aufgaben kann ich in der Regel gut lösen. | 5-point scales from "does not apply at all" to "applies completely" |

(continued on next page)

**Table A3 (continued)**

| | Wording (English translation)<br>(1) | Wording (German original)<br>(2) | Answer categories<br>(3) |
|---|---|---|---|
| *Labor-market orientation index* | *Equally weighted average of two components: wants apprenticeship after school; knows future career* | | |
| Wants apprenticeship after school | Follow-up 1: answer "Apprenticeship" to "What would you like to do after finishing school?" | Was möchtest du nach deinem gewünschten Schulabschluss machen? Ausbildung. | Apprenticeship; university; directly entering a job; something else; don't know yet |
| | Follow-up 2: answer "Apprenticeship" to "Do you already have concrete plans for what you will do after school?" | Hast du schon konkrete Pläne dafür, was du nach der Schule machen wirst? Ausbildung | Apprenticeship; higher-track secondary school (Abitur); vocational school; pre-employment training; work; something else; don't know |
| | Follow-up 3: answer "Apprenticeship" to "Are you aiming for one or more of the following degrees in the future" | Strebst du in der Zukunft einen oder mehrere der folgenden Abschlüsse an? Ausbildung | Apprenticeship; higher-track secondary school (Abitur); vocational school; technical school; university of applied science; university |
| Knows future career | Agreement to item: I already know exactly which occupation I want to work in later in life. | Ich weiß schon genau, was ich später mal beruflich machen will. | 4-point scale from "do not agree at all" to "agree" |

(continued on next page)

**Table A3 (continued)**

| | Wording (English translation)<br>(1) | Wording (German original)<br>(2) | Answer categories<br>(3) |
|---|---|---|---|
| Early-career realization | Follow-up 3: What are you currently doing? | Was machst du gerade? | Apprenticeship; school; pre-employment training; internship; gap year to do voluntary work; work; unemployed; something else |
| *Satisfaction with current situation* | | | |
| Satisfied | Follow-up 3: How satisfied are you with your current situation? | Wie zufrieden bist du damit, … zu machen? | 4-point scale from “not satisfied” to “very satisfied” |
| No desire to change | Follow-up 3: Would you like to change something about your situation? | Würdest du gerne etwas an deiner Situation ändern? | Yes; no |
| **Covariates (baseline survey)** | | | |
| Male | Answer “male” to “Are you male or female?” | Bist du männlich oder weiblich? | Male; female |
| Age | Based on “When were you born?” | Wann bist du geboren? | Day, month, and year of birth |
| Migrant | Adolescent or at least one parent not born in Germany. | In welchem Land bist du geboren?<br>In welchem Land ist deine Mutter geboren?<br>In welchem Land ist dein Vater geboren? | Germany; other country (name) |
| Received paid private teaching | Answer “Yes” to “Did you get paid private teaching in the last semester of school?” | Hast du im letzten Schulhalbjahr bezahlten Nachhilfeunterricht bekommen? | Yes; no |
| Parental homework support | Do your parents (mother and/or father) support you with your homework and learning for school? | Unterstützen dich deine Eltern (Mutter und/oder Vater) bei den Hausaufgaben und beim Lernen für die Schule? | 4-point scale: not at all; rather little; rather strong; very strong |

(continued on next page)

**Table A3 (continued)**

| | Wording (English translation) (1) | Wording (German original) (2) | Answer categories (3) |
|---|---|---|---|
| *Big-5 personality traits* | *Personality inventory according to 10-Item Big-5 Inventory (Rammstedt (2007); Rammstedt and John (2007))* | | |
| Conscientiousness | Agreement to two items: I am someone who tends to be lazy (reversed); I am someone who does a thorough job. | Ich bin bequem, neige zur Faulheit; Ich erledige Aufgaben gründlich. | 5-point scales from "does not apply at all" to "applies completely" |
| Neuroticism | Agreement to two items: I am someone who is relaxed, handles stress well (reversed); I am someone who gets nervous easily. | Ich bin entspannt, lasse mich durch Stress nicht aus der Ruhe bringen; Ich werde leicht nervös und unsicher. | 5-point scales from "does not apply at all" to "applies completely" |
| Openness | Agreement to two items: I am someone who has few artistic interests (reversed); I am someone who has a vivid imagination/fantasy. | Ich habe nur wenig künstlerisches Interesse; Ich habe eine aktive Vorstellungskraft, bin fantasievoll. | 5-point scales from "does not apply at all" to "applies completely" |
| Extraversion | Agreement to two items: I am someone who is reserved (reversed); I am someone who is outgoing, sociable. | Ich bin eher zurückhaltend, reserviert; Ich gehe aus mir heraus, bin gesellig. | 5-point scales from "does not apply at all" to "applies completely" |
| Agreeableness | Agreement to two items: I am someone who is generally trusting; I am someone who tends to find fault with others (reversed). | Ich schenke anderen leicht Vertrauen, glaube an das Gute im Menschen; Ich neige dazu, andere zu kritisieren. | 5-point scales from "does not apply at all" to "applies completely" |
| **Further matching and balancing variables (baseline)** | | | |
| >25 books at home | Answer more than 25 to "Approximately how many books are there in your home?" | Wie viele Bücher gibt es bei dir zuhause ungefähr? Antworten: genug, um mehrere Regalbretter zu füllen (26 bis 100 Bücher), genug, um ein kleines Regal zu füllen (101 bis 200 Bücher), genug, um ein großes Regal zu füllen (201 bis 500 Bücher), genug, um eine Regalwand zu füllen (mehr als 500 Bücher). | Books: 0-10; 11-25; 26-100; 101-200; 201-500; more than 500 |

Notes: All indices are constructed as equally weighted average of the *z*-scores of the included items; calculation of each *z*-score subtracts the score's control-group mean and divides by the control-group standard deviation (Kling, Liebman, and Katz (2007)).

**Table A4: Overall Effect of the Mentoring Program: Splitting the Sample by SES**

| | SES | | SES pair | |
|---|---|---|---|---|
| | Low-SES | Higher-SES | Low-SES | Higher-SES |
| | (1) | (2) | (3) | (4) |
| **A. Without randomization-pair fixed effects** | | | | |
| Treatment | 0.634*** | -0.124 | 0.799*** | -0.178 |
| | *(0.158)* | *(0.128)* | *(0.188)* | *(0.158)* |
| Outcome in $t_0$ | 0.580*** | 0.417*** | 0.623*** | 0.439*** |
| | *(0.075)* | *(0.082)* | *(0.093)* | *(0.112)* |
| Randomization-pair fixed effects | No | No | No | No |
| Covariates | Yes | Yes | Yes | Yes |
| Observations | 138 | 166 | 82 | 106 |
| $R^2$ | 0.527 | 0.261 | 0.677 | 0.323 |
| **B. With randomization-pair fixed effects** | | | | |
| Treatment | 0.725*** | -0.231 | 0.709*** | -0.194 |
| | *(0.185)* | *(0.182)* | *(0.153)* | *(0.163)* |
| Outcome in $t_0$ | 0.673*** | 0.464** | 0.665*** | 0.453** |
| | *(0.148)* | *(0.178)* | *(0.124)* | *(0.189)* |
| Randomization-pair fixed effects | Yes | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes | Yes |
| Observations | 138 | 166 | 82 | 106 |
| $R^2$ | 0.951 | 0.801 | 0.932 | 0.701 |

Notes: Table shows ITT effects of the mentoring program on the index of labor-market prospects one year after program start. Columns 1 and 2 split the sample by individual SES status. Columns 3 and 4 split the sample by pairs in which both adolescents either have a low-SES or a higher-SES background; i.e., mixed-SES pairs are dropped from the analysis. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in $t_0$ are included. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

**Table A5: Effect of the Mentoring Program by Gender**

| | (1) | (2) | (3) |
|---|---|---|---|
| Treatment | 0.644*** | 0.212 | 0.655*** |
| | *(0.149)* | *(0.171)* | *(0.196)* |
| Treatment x Higher-SES | -0.865*** | | -0.863*** |
| | *(0.227)* | | *(0.228)* |
| Treatment x Female | | -0.104 | -0.020 |
| | | *(0.215)* | *(0.204)* |
| Higher-SES | 0.286 | | 0.288 |
| | *(0.192)* | | *(0.193)* |
| Female | 0.189 | 0.257 | 0.199 |
| | *(0.221)* | *(0.241)* | *(0.245)* |
| Outcome in $t_0$ | 0.446*** | 0.440*** | 0.446*** |
| | *(0.076)* | *(0.083)* | *(0.077)* |
| Randomization-pair fixed effects | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes |
| Observations | 304 | 304 | 304 |
| $R^2$ | 0.758 | 0.731 | 0.758 |
| Treatment effect for Higher-SES | -0.221 | | -0.208 |
| | *(0.136)* | | *(0.203)* |
| Treatment effect for Females | | 0.108 | |
| | | *(0.108)* | |
| Treatment effect for Females, low-SES | | | 0.634*** |
| | | | *(0.164)* |

Notes: Table shows ITT effects of the mentoring program on the index of labor-market prospects one year after program start. Covariates are from the baseline survey and include: age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

**Table A6: Effect of the Mentoring Program on the Three Components of the Index of Labor-Market Prospects: Correction for Multiple Hypotheses Testing**

| | Math grade | Patience and social skills | Labor-market orientation |
|---|---|---|---|
| | (1) | (2) | (3) |
| Treatment | 0.309 | 0.468 | 0.443 |
| Standard *p*-value | *0.033* | *0.005* | *0.010* |
| List-Shaikh-Xu *p*-value | *0.037* | *0.028* | *0.019* |
| Westfall-Young *p*-value | *0.032* | *0.016* | *0.016* |
| Treatment x Higher-SES | -0.493 | -0.593 | -0.544 |
| Standard *p*-value | *0.026* | *0.024* | *0.048* |
| List-Shaikh-Xu *p*-value | *0.061* | *0.063* | *0.054* |
| Westfall-Young *p*-value | *0.074* | *0.074* | *0.074* |
| Higher-SES indicator | Yes | Yes | Yes |
| Outcome in $t_0$ | Yes | Yes | Yes |
| Randomization-pair fixed effects | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes |
| Observations | 294 | 291 | 291 |

Notes: Table shows ITT effects of the mentoring program on administrative math grade (reversed), patience and social skills index, and labor-market orientation index one year after program start. The three columns replicate the specifications in columns 1, 3, and 6 of Table 3. In addition to the standard *p*-values based on robust standard errors, the table reports *p*-values robust to multiple hypothesis testing (family-wise error rates) using the bootstrap resampling techniques by List, Shaikh, and Xu (2019) and Westfall and Young (1993), respectively. Bootstraps are adjusted to account for the pair structure in the data, i.e., the sample drawn during each replication is a bootstrap sample of pairs. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in $t_0$ are included.

**Table A7: Effect of the Mentoring Program on Measures of Social Capital**

| | Volunteer | Friends | | School-related activities | | |
|---|---|---|---|---|---|---|
| | | Number of friends | Number of meetings | All activities | High-stakes activities | Low-stakes activities |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment | 0.029 | 0.015 | 0.094 | -0.110 | 0.009 | -0.157 |
| | *(0.086)* | *(0.143)* | *(0.243)* | *(0.215)* | *(0.195)* | *(0.232)* |
| | *[0.723]* | *[0.931]* | *[0.680]* | *[0.611]* | *[0.960]* | *[0.477]* |
| Treatment x Higher-SES | -0.020 | -0.292 | -0.119 | -0.096 | -0.090 | -0.054 |
| | *(0.128)* | *(0.276)* | *(0.311)* | *(0.313)* | *(0.295)* | *(0.336)* |
| | *[0.873]* | *[0.268]* | *[0.705]* | *[0.765]* | *[0.786]* | *[0.871]* |
| Higher-SES | 0.038 | 0.048 | 0.213 | 0.064 | -0.054 | 0.134 |
| | *(0.093)* | *(0.247)* | *(0.287)* | *(0.232)* | *(0.227)* | *(0.239)* |
| Outcome in $t_0$ | 0.556$^{***}$ | 0.234$^{**}$ | 0.343$^{***}$ | – | – | – |
| | *(0.075)* | *(0.098)* | *(0.110)* | | | |
| Randomization-pair fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 290 | 284 | 287 | 290 | 290 | 290 |
| $R^2$ | 0.762 | 0.577 | 0.600 | 0.630 | 0.642 | 0.558 |
| Treatment effect for Higher-SES | 0.009 | -0.276 | -0.024 | -0.206 | -0.081 | -0.210 |
| | *(0.074)* | *(0.186)* | *(0.158)* | *(0.176)* | *(0.177)* | *(0.189)* |

Notes: Table shows ITT effects of the mentoring program on social capital domains (elicited one year after program start) indicated in the column header. *Volunteer:* Are you volunteering at least once a week in a club or association (e.g., sports clubs, youth clubs, voluntary fire brigade, supporters club, political parties, musical and artistic groups, etc.)? *Number of friends:* How many friends do you regularly meet in your private time, i.e., outside of school time. Number of friends is standardized. *Number of meetings:* How many times do you meet with friends in your private time, i.e., outside of school time. Number of meeting days in a regular week is standardized. *School-related activities:* Average index of the following school activities (represented by a dummy variable that is one if true, and zero otherwise): acting as class representative, working as peer mediator, acting as school representative, working for the school magazine, volunteering as school nurse, participating in the school music ensemble, participating in the school theater group, and participating in other school activity. *High-stakes activities:* Only indicates more high-stakes activities: acting as class representative, working as peer mediator, and acting as school representative. *Low-stakes activities:* Collects the remaining activities. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in $t_0$ are included. Robust standard errors in parentheses. Randomization inference (RI) *p*-values in square brackets, obtained from RI with 1,000 permutations, assigning the treatment status randomly within randomization pairs. Significance levels: $^{***}$ p < 0.01, $^{**}$ p < 0.05, $^{*}$ p < 0.1.

**Table A8: Effect of the Mentoring Program on Plans after School**

| | Apprenticeship (1) | University (2) | Don't know (3) | Direct job (4) | Other (5) |
|---|---|---|---|---|---|
| Treatment | 0.260*** | -0.104 | -0.109 | 0.013 | -0.017 |
| | *(0.081)* | *(0.088)* | *(0.086)* | *(0.044)* | *(0.048)* |
| Treatment x Higher-SES | -0.353** | 0.213 | 0.070 | -0.045 | 0.025 |
| | *(0.137)* | *(0.130)* | *(0.131)* | *(0.059)* | *(0.065)* |
| Higher-SES | 0.233** | -0.095 | -0.078 | 0.062* | 0.008 |
| | *(0.107)* | *(0.102)* | *(0.112)* | *(0.037)* | *(0.051)* |
| Outcome in $t_0$ | 0.486*** | 0.466*** | 0.319*** | -0.029 | 0.171 |
| | *(0.084)* | *(0.094)* | *(0.095)* | *(0.039)* | *(0.179)* |
| Randomization-pair fixed effects | Yes | Yes | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes | Yes | Yes |
| Observations | 290 | 290 | 290 | 290 | 290 |
| $R^2$ | 0.675 | 0.644 | 0.587 | 0.579 | 0.538 |
| Treatment effect for Higher-SES | -0.094 | 0.109 | -0.039 | -0.031 | 0.009 |
| | *(0.092)* | *(0.082)* | *(0.087)* | *(0.028)* | *(0.036)* |

Notes: Table shows ITT effects of the mentoring program on respondents' wishes for their plans after leaving school, elicited one year after program start. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in $t_0$ are included. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

**Table A9: Effect of the Mentoring Program by Migrant Status: First- and Second-Generation Migrants**

| | Outcome index | | Math grade | | Patience and social skills index | | Labor-market orientation index | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Treatment | 0.290** | 0.730** | 0.123 | 0.385 | 0.068 | 0.653* | 0.272* | 0.285 |
| | *(0.122)* | *(0.290)* | *(0.123)* | *(0.263)* | *(0.159)* | *(0.388)* | *(0.142)* | *(0.317)* |
| Treatment x Non-migrant | -0.328 | -0.751** | -0.198 | -0.452 | 0.143 | -0.418 | -0.331 | -0.353 |
| | *(0.230)* | *(0.327)* | *(0.215)* | *(0.301)* | *(0.299)* | *(0.449)* | *(0.260)* | *(0.372)* |
| Non-migrant | 0.323* | 0.631** | 0.234 | 0.368 | -0.080 | -0.050 | 0.264 | 0.639** |
| | *(0.193)* | *(0.292)* | *(0.176)* | *(0.266)* | *(0.217)* | *(0.378)* | *(0.230)* | *(0.296)* |
| Treatment x Second-gen. migrant | | -0.596 | | -0.339 | | -0.743 | | -0.051 |
| | | *(0.382)* | | *(0.332)* | | *(0.467)* | | *(0.409)* |
| Second-generation migrant | | 0.417 | | 0.183 | | 0.096 | | 0.439 |
| | | *(0.283)* | | *(0.253)* | | *(0.384)* | | *(0.276)* |
| Outcome in $t_0$ | 0.449*** | 0.439*** | 0.520*** | 0.517*** | 0.259** | 0.267*** | 0.375*** | 0.382*** |
| | *(0.083)* | *(0.082)* | *(0.103)* | *(0.104)* | *(0.101)* | *(0.102)* | *(0.090)* | *(0.088)* |
| Randomization-pair fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 304 | 304 | 294 | 294 | 291 | 291 | 291 | 291 |
| $R^2$ | 0.734 | 0.739 | 0.769 | 0.770 | 0.683 | 0.691 | 0.692 | 0.698 |
| Treatment effect for non-migrants | -0.038 | -0.021 | -0.075 | -0.067 | 0.211 | 0.235 | -0.059 | -0.068 |
| | *(0.170)* | *(0.174)* | *(0.156)* | *(0.157)* | *(0.208)* | *(0.212)* | *(0.190)* | *(0.192)* |
| Treatment effect for second-generation migrants | | 0.134 | | 0.046 | | -0.090 | | 0.234 |
| | | *(0.166)* | | *(0.156)* | | *(0.189)* | | *(0.186)* |
| Migrant gap | 0.224 | | 0.314 | | -0.075 | | 0.162 | |
| First-generation migrant gap | 0.125 | | 0.428 | | -0.326 | | 0.130 | |
| Second-generation migrant gap | 0.256 | | 0.279 | | 0.006 | | 0.172 | |

Notes: Table shows ITT effects of the mentoring program on the outcome indicated in the column header, elicited one year after program start. *Non-migrant* indicates that an individual and both of his/her parents were born in Germany (i.e., *migrants* are first-generation or second-generation migrants). Covariates are from the baseline survey and include: gender, age, received paid private teaching, parental homework support, and Big-5 personality traits. The migrant gap is calculated as the coefficient on non-migrant in a regression of the respective outcome on an indicator for non-migrant. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

**Table A10: Evidence on the Mentoring Relationships**

| | All | Low-SES | Higher-SES | Difference | |
|---|---|---|---|---|---|
| | | | | (2)-(3) | *p*-value |
| | (1) | (2) | (3) | (4) | (5) |
| **A. Initiation and continuation of relationship** | | | | | |
| Mentee has met mentor at least once | 0.86 | 0.83 | 0.89 | -0.05 | 0.371 |
| Mentoring relationship still exists | 0.63 | 0.56 | 0.69 | -0.13 | 0.105 |
| Mentoring relationship still exists (conditional on mentor/mentee ever met) | 0.73 | 0.67 | 0.77 | -0.09 | 0.246 |
| **B. Meeting frequency and duration** | | | | | |
| Meet at least once per month (in person) | 0.50 | 0.50 | 0.50 | 0.00 | 1.00 |
| Meet at least once per month (all channels) | 0.61 | 0.57 | 0.65 | -0.08 | 0.364 |
| Duration of meetings (hours) | 3.13 | 2.95 | 3.23 | -0.32 | 0.455 |
| **C. Topics discussed during meetings** | | | | | |
| School | 0.66 | 0.62 | 0.68 | -0.06 | 0.429 |
| Leisure activities | 0.57 | 0.49 | 0.63 | -0.14 | 0.089 |
| Future in general | 0.57 | 0.59 | 0.55 | 0.04 | 0.645 |
| Occupational and educational future | 0.50 | 0.52 | 0.49 | 0.04 | 0.670 |
| Personal issues | 0.49 | 0.49 | 0.49 | 0.00 | 0.960 |
| Family issues | 0.25 | 0.29 | 0.22 | 0.07 | 0.369 |
| Other topics | 0.13 | 0.10 | 0.16 | -0.06 | 0.253 |
| Don’t know | 0.20 | 0.24 | 0.17 | 0.07 | 0.326 |
| Mentee can decide what is done in meetings | 0.62 | 0.58 | 0.64 | -0.06 | 0.460 |
| **D. Qualitative factors of relationship** | | | | | |
| Mentee better at school because of mentor | 0.20 | 0.26 | 0.16 | 0.10 | 0.173 |
| Mentor helped solve non-school-related problems | 0.30 | 0.36 | 0.26 | 0.10 | 0.225 |
| Mentor is role model | 0.27 | 0.31 | 0.24 | 0.07 | 0.366 |
| Parents like that their child has mentor | 0.54 | 0.44 | 0.62 | -0.18 | 0.031 |
| Friends support mentee having a mentor | 0.26 | 0.26 | 0.26 | -0.00 | 0.987 |
| Mentee had a say in which mentor he/she got | 0.47 | 0.41 | 0.51 | -0.11 | 0.219 |
| Mentee and mentor are good friends | 0.49 | 0.52 | 0.48 | 0.04 | 0.630 |
| Mentee satisfied with mentoring relationship | 0.59 | 0.53 | 0.63 | -0.09 | 0.271 |

Notes: Table shows group means of variables characterizing the mentoring relationships, elicited in the treatment group one year after program start. Sample: column 1: all respondents ($n$=153); column 2: low-SES respondents ($n$=66); column 3: higher-SES respondents ($n$=87). Column 5 shows the $p$-value of the coefficient on the higher-SES indicator in a regression of the specific variable on a higher-SES indicator.

**Table A11: Effect of the Mentoring Program on School Performance**

| | Math achievement | | | Graduation | |
|---|---|---|---|---|---|
| | ~0.75 years | ~1.5 years after program start | ~2 years | Any certificate | Upper certificate |
| | (1) | (2) | (3) | (4) | (5) |
| Treatment | 0.584*** | 0.377* | 0.353* | 0.033 | 0.086 |
| | *(0.209)* | *(0.204)* | *(0.209)* | *(0.073)* | *(0.090)* |
| Treatment x Higher-SES | -0.836*** | -0.567** | -0.566** | -0.048 | -0.085 |
| | *(0.277)* | *(0.267)* | *(0.273)* | *(0.105)* | *(0.126)* |
| Higher-SES | 0.274 | 0.172 | 0.103 | 0.069 | 0.030 |
| | *(0.234)* | *(0.250)* | *(0.264)* | *(0.061)* | *(0.106)* |
| Outcome in $t_0$ | 0.536*** | 0.711*** | 0.669*** | – | – |
| | *(0.119)* | *(0.125)* | *(0.155)* | | |
| Randomization-pair fixed effects | Yes | Yes | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes | Yes | Yes |
| Observations | 176 | 176 | 177 | 300 | 300 |
| $R^2$ | 0.781 | 0.794 | 0.778 | 0.584 | 0.676 |
| Treatment effect for Higher-SES | -0.252 | -0.190 | -0.213 | -0.017 | 0.000 |
| | *(0.167)* | *(0.190)* | *(0.196)* | *(0.054)* | *(0.074)* |
| SES gap | 0.446 | 0.273 | 0.339 | 0.066 | 0.224 |

Notes: Table shows ITT effects of the mentoring program on the standardized math grade (columns 1-3) and on graduation (columns 4 and 5). Sample in columns 1-2 is restricted to individuals in randomization pairs who are in the same class across all periods. Graduation is a dummy variable taking a value of one if the adolescent has graduated, and zero otherwise. Column 5: adolescent has graduated with at least a middle-school graduation certificate (*Mittlerer Schulabschluss*). Covariates are from the baseline survey and include: labor-market prospects index in columns 4 and 5, gender, age, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in $t_0$ are included. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

**Table A12: Effect of the Mentoring Program on Career Vision**

| | Has desired occupation | Has desired or apprenticeship occupation |
|---|---|---|
| | (1) | (2) |
| Treatment | 0.186* | 0.264** |
| | *(0.096)* | *(0.105)* |
| Treatment x Higher-SES | -0.296* | -0.347** |
| | *(0.156)* | *(0.160)* |
| Higher-SES | 0.073 | 0.137 |
| | *(0.127)* | *(0.131)* |
| Randomization-pair fixed effects | Yes | Yes |
| Covariates | Yes | Yes |
| Observations | 272 | 272 |
| $R^2$ | 0.658 | 0.623 |
| Treatment effect for Higher-SES | -0.110 | -0.083 |
| | *(0.103)* | *(0.103)* |
| SES gap | -0.003 | -0.007 |

Notes: Table shows ITT effects of the mentoring program on the propensity that adolescents reported an actual or desired occupation, elicited three years after program start. We elicit the desired occupation using the following question: *Do you already know which occupation you want to take up?* Answer categories were: yes, with great certainty; yes, with some certainty; no, still open. Only respondents who answered yes (i.e., the first two categories) were asked about the specific occupation in which they would like to work. Respondents provided up to three desired occupations (159 respondents provided only one desired occupation; 16 respondents provided two desired occupations, 2 respondents provided three desired occupations). In column 1, the dependent variable is a binary variable, taking a value of one if adolescents reported at least one desired occupation, and zero otherwise. In column 2, we also take into account if respondents already have an occupation: the dependent variable is a binary variable, taking a value of one if adolescents reported a desired occupation or already have an apprenticeship occupation, and zero otherwise. Covariates are from the baseline survey and include: labor-market orientation index, gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in $t_0$ are included. Additional controls include fixed effects for surveys that were carried out during the first COVID-19 wave (March 2020 to May 2020) or the second COVID-19 wave (October 2020 to February 2021) in Germany. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

**Table A13: Effect of the Mentoring Program on Successful Completion and Expected Labor-Market Returns**

| | Probability of successful completion | | Expected returns | |
|---|---|---|---|---|
| | University | Apprenticeship | University | Apprenticeship |
| | (1) | (2) | (3) | (4) |
| Treatment | -13.418* | -1.598 | -0.007 | -0.053 |
| | *(7.630)* | *(4.603)* | *(0.125)* | *(0.067)* |
| Treatment x Higher-SES | 7.081 | -5.189 | -0.048 | 0.060 |
| | *(10.605)* | *(7.649)* | *(0.162)* | *(0.097)* |
| Higher-SES | 2.934 | 0.961 | 0.079 | 0.047 |
| | *(8.071)* | *(5.398)* | *(0.123)* | *(0.087)* |
| Randomization-pair fixed effects | Yes | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes | Yes |
| Observations | 271 | 271 | 268 | 269 |
| $R^2$ | 0.628 | 0.660 | 0.640 | 0.619 |
| Treatment effect for Higher-SES | -6.338 | -6.787 | -0.056 | 0.007 |
| | *(5.993)* | *(4.682)* | *(0.071)* | *(0.058)* |
| SES gap | 5.457 | -0.250 | 0.038 | 0.010 |

Notes: Table shows ITT effects of the mentoring program on successful completion of and expected labor-market returns to university education and apprenticeship training, respectively, elicited three years after program start. Columns 1 and 2: self-assessed likelihood (in percent) that respondents successfully complete university education or apprenticeship training, respectively, if they would have started it. Columns 3 and 4: respondents' estimate of expected net monthly earnings when having completed university education or apprenticeship training, respectively. To exclude outliers, we drop the top 0.5 percent of reported earnings. Earnings are expressed in logarithms. Covariates are from the baseline survey and include: labor-market orientation index, gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in $t_0$ are included. Additional controls include fixed effects for surveys that were carried out during the first COVID-19 wave (March 2020 to May 2020) or the second COVID-19 wave (October 2020 to February 2021) in Germany. Robust standard errors in parentheses. Significance levels: $^{***}$ $p < 0.01$, $^{**}$ $p < 0.05$, $^{*}$ $p < 0.1$.

## Appendix B: Survey Frame: Sites, Cohorts, and Timing

This appendix describes the selection criteria for sites to participate in the RCT (Appendix B.1) and the two-cohort sampling frame (Appendix B.2).

### B.1 Selection of Participating Sites

Among the 42 sites served by the mentoring program in Germany, we aimed to approach locations for participation in regions that are representative for the target population of the mentoring program. In particular, these included large cities (e.g., Berlin, Hamburg, Cologne) and agglomeration areas (Rhine-Ruhr area) with a high share of disadvantaged youths. Moreover, we approached sites that were already established before the start of the RCT (i.e., operating for at least two years) and that were likely to reach the oversubscription needed for the randomization. By applying these site selection criteria, we avoided cream skimming by the mentoring program (i.e., selection of sites that are expected to produce the highest benefits for the adolescents; see Heckman (2020)).

In each site satisfying the selection criteria, we approached the university-student officials of the respective university society to ask for their cooperation. Officials from the holding helped with establishing the contacts and were personally present in several meetings. Eventually, all contacted sites agreed to cooperate. Together with officials from the university society and the holding, we then personally approached the principal of each cooperating school to get permission to conduct the surveys in the schools during class hours to maximize participation. All schools agreed to participate in the surveys.

Carrying out the surveys at school also required receiving the approval by the respective states' school administrative bodies. We received approval from all but one state where we intended to survey participants.[1] The six states are: Baden-Wurttemberg (for Mannheim), Berlin (for Berlin), Hamburg (for Hamburg), Schleswig-Holstein (for Luebeck), Lower Saxony (for Lueneburg), and Saxony (for Chemnitz and Leipzig). Schools in the state of North-Rhine Westphalia (for Aachen, Bonn, Cologne, Duisburg, and Essen) are allowed to approve requests from researchers on their own discretion.

[1] Bavaria refused to provide permission to conduct the study in schools in their federal state due to general ethical concerns to conduct randomized trials (although the schools had already agreed to participate).

## B.2 Two-Cohort Sampling Frame

In our pre-analysis plan (contained in the grant application registered with the funding foundations on May 12, 2015), we envisaged a two-cohort sampling procedure to provide a sufficiently large sample to reliably estimate treatment effects. Figure 1 depicts the timeline of the baseline and follow-up surveys in the two cohorts. Table B1 shows the dates and sample sizes by cohort and mentoring site for the baseline survey.

The first cohort includes youths in 17 schools in nine cities (organized in eight mentoring sites). The survey period began with a couple of pilot studies in Aachen in November 2015 and in Duisburg in June 2016 which were used to test the main features of the evaluation, i.e., communication with principals, teachers, and mentoring society officials, collection of baseline data in the applicant survey, randomization procedure, and dissemination of assignment decisions. Because we had already tested the questionnaires extensively prior to the pilot studies and only few minor adaptations in the design were necessary after the pilot studies, we decided to include the pilot data in the main evaluation.[2]

Further data collection in the first cohort proceeded in three phases, because start dates of new mentoring cohorts differed between mentoring sites. In October and November 2016, we collected baseline data (in chronological order) in Mannheim, Cologne, Essen, Aachen, Berlin, and Luebeck. In January 2017, we collected data in Hamburg. In May 2017, we collected data in Lueneburg and two other schools in Berlin.

The second cohort, which started about one year after the first cohort, includes youths in 21 schools in ten cities/mentoring sites. The second cohort comprised seven sites already included in the first cohort and three new sites which promised reasonably good oversubscription. Specifically, between October and December 2017, we collected baseline data in Leipzig (new site), Bonn (new site), Berlin, Cologne, Chemnitz (new site), Lueneburg, Aachen, Luebeck, Essen, and Hamburg. In May 2018, we collected data in another school in Berlin.

Table B1 provides information on the total and randomized samples in each site. Table B2 shows that adolescents who could not be included in the randomization are similar to those in the randomized sample. The only differences that are statistically significant at the 5 percent level are in patience and the shares of missing survey observations on math grades.

[2] Results are robust to excluding the pilot cohort (not shown).

**Table B1: Observations in the Baseline Survey by Mentoring Site and Cohort**

| Cohort | Site/city | School ID | Survey period | | Total sample | | Randomized sample | |
|---|---|---|---|---|---|---|---|---|
| | | | Month | Year | Control | Treatment | Control | Treatment |
| 1 | Aachen[a] | 1 | 11 | 2015 | 14 | 15 | 14 | 14 |
| 1 | Aachen | 1 | 11 | 2016 | 15 | 14 | 14 | 13 |
| 1 | Berlin | 1 | 11 | 2016 | 3 | 4 | 3 | 4 |
| 1 | Berlin | 2 | 5 | 2017 | 8 | 7 | 8 | 7 |
| 1 | Berlin | 3 | 5 | 2017 | 6 | 8 | 6 | 6 |
| 1 | Cologne | 1 | 11 | 2016 | 7 | 7 | 7 | 7 |
| 1 | Cologne | 2 | 11 | 2016 | 6 | 6 | 6 | 6 |
| 1 | Cologne | 3 | 11 | 2016 | 4 | 5 | 4 | 5 |
| 1 | Duisburg[a,b] | 1 | 6 | 2016 | 6 | 7 | 6 | 7 |
| 1 | Essen[a,b] | 1 | 11 | 2016 | 5 | 5 | 4 | 4 |
| 1 | Hamburg | 1 | 1 | 2017 | 5 | 4 | 5 | 4 |
| 1 | Hamburg | 2 | 1 | 2017 | 7 | 6 | 7 | 6 |
| 1 | Hamburg | 3 | 1 | 2017 | 2 | 2 | 2 | 2 |
| 1 | Hamburg | 4 | 1 | 2017 | 1 | 6 | — | — |
| 1 | Luebeck | 1 | 11 | 2016 | 20 | 13 | 13 | 13 |
| 1 | Luebeck | 2 | 11 | 2016 | 8 | 12 | 8 | 8 |
| 1 | Lueneburg | 1 | 5 | 2017 | 0 | 6 | — | — |
| 1 | Mannheim | 1 | 10 | 2016 | 4 | 6 | 4 | 4 |
| 2 | Aachen | 1 | 11 | 2017 | 0 | 11 | — | — |
| 2 | Aachen | 2 | 11 | 2017 | 0 | 2 | — | — |
| 2 | Berlin | 1 | 11 | 2017 | 5 | 7 | 5 | 5 |
| 2 | Berlin | 2 | 5 | 2018 | 0 | 5 | — | — |
| 2 | Berlin | 3 | 5 | 2018 | 8 | 7 | 4 | 4 |
| 2 | Berlin | 4 | 11 | 2017 | 0 | 6 | — | — |
| 2 | Bonn | 1 | 11 | 2017 | 0 | 6 | — | — |
| 2 | Chemnitz | 1 | 11 | 2017 | 0 | 6 | — | — |
| 2 | Chemnitz | 2 | 11 | 2017 | 0 | 4 | — | — |
| 2 | Cologne | 1 | 11 | 2017 | 8 | 7 | 8 | 7 |
| 2 | Cologne | 2 | 11 | 2017 | 4 | 4 | 4 | 3 |
| 2 | Essen[a] | 1 | 12 | 2017 | 0 | 6 | — | — |
| 2 | Hamburg | 1 | 12 | 2017 | 2 | 5 | 2 | 2 |
| 2 | Hamburg | 2 | 12 | 2017 | 1 | 4 | 1 | 1 |
| 2 | Hamburg | 3 | 12 | 2017 | 0 | 4 | — | — |
| 2 | Hamburg | 4 | 12 | 2017 | 4 | 5 | 4 | 5 |
| 2 | Leipzig | 1 | 10 | 2017 | 2 | 7 | 2 | 2 |
| 2 | Leipzig | 2 | 10 | 2017 | 1 | 3 | 1 | 1 |
| 2 | Luebeck | 1 | 12 | 2017 | 5 | 19 | 5 | 5 |
| 2 | Luebeck | 2 | 11 | 2017 | 0 | 10 | — | — |
| 2 | Lueneburg | 2 | 11 | 2017 | 8 | 12 | 8 | 8 |
| | | | | | 169 | 273 | 155 | 153 |
| | | | | | 442 | | 308 | |

Notes: Table shows dates and sample sizes of the baseline survey for each site and cohort. [a] Pilot studies. [b] Duisburg and Essen belong to the same mentoring site. "—": randomization was not possible due to lack of oversubscription.

**Table B2: Comparison of Randomized and Non-Randomized Samples**

| | Sample | | | Difference | | |
|---|---|---|---|---|---|---|
| | Total *N=442* | Randomized *N=308* | Non-rand. *N=134* | (2)-(3) | *p*-value | Obs. |
| | (3) | (1) | (2) | (4) | (5) | (6) |
| **A. Outcome variables at baseline** | | | | | | |
| Overall index | -0.05 | -0.04 | -0.05 | 0.01 | 0.936 | 442 |
| *Components* | | | | | | |
| Math grade (administrative) | -0.03 | 0.01 | -0.12 | 0.13 | 0.308 | 311 |
| Math grade (admin.) missing d. | 0.30 | 0.29 | 0.31 | -0.01 | 0.773 | 442 |
| Patience and social skills index | -0.07 | -0.04 | -0.16 | 0.12 | 0.279 | 442 |
| Labor-market orientation index | 0.02 | -0.04 | 0.15 | -0.20 | 0.064 | 441 |
| **B. Components of outcome variables at baseline** | | | | | | |
| *Patience and social skills index* | | | | | | |
| Patience | -0.08 | -0.01 | -0.24 | 0.23 | 0.044 | 441 |
| Social skills index | -0.03 | -0.05 | 0.00 | -0.05 | 0.643 | 442 |
| *Components* | | | | | | |
| Prosociality | 0.06 | 0.01 | 0.17 | -0.16 | 0.082 | 442 |
| Trust | -0.03 | -0.02 | -0.06 | 0.03 | 0.754 | 438 |
| Self-efficacy | -0.09 | -0.08 | -0.11 | 0.03 | 0.767 | 441 |
| *Labor-market orientation index* | | | | | | |
| Wants apprenticeship after school | 0.40 | 0.37 | 0.45 | -0.08 | 0.109 | 400 |
| Knows future career | -0.04 | -0.08 | 0.06 | -0.14 | 0.156 | 439 |
| **C. Matching and balancing variables for randomization at baseline** | | | | | | |
| Male | 0.43 | 0.44 | 0.43 | 0.00 | 0.965 | 442 |
| Age | 14.00 | 13.98 | 14.04 | -0.06 | 0.504 | 442 |
| Migrant | 0.57 | 0.58 | 0.57 | 0.01 | 0.834 | 442 |
| Books at home | 1.72 | 1.70 | 1.76 | -0.06 | 0.409 | 442 |
| Math grade (survey) | 1.74 | 1.72 | 1.77 | -0.05 | 0.580 | 361 |
| Math grade (survey) missing d. | 0.18 | 0.15 | 0.25 | -0.10 | 0.019 | 442 |
| German grade (survey) | 1.74 | 1.72 | 1.79 | -0.07 | 0.368 | 359 |
| German grade (survey) missing d. | 0.19 | 0.16 | 0.25 | -0.08 | 0.051 | 442 |
| English grade (survey) | 1.80 | 1.81 | 1.76 | 0.05 | 0.516 | 359 |
| English grade (survey) missing d. | 0.19 | 0.16 | 0.25 | -0.08 | 0.051 | 442 |
| Received paid private teaching | 0.20 | 0.19 | 0.20 | 0.00 | 0.987 | 441 |
| Parental homework support | 2.75 | 2.76 | 2.74 | 0.02 | 0.868 | 441 |
| Big-5: Conscientiousness | 3.28 | 3.31 | 3.22 | 0.09 | 0.289 | 442 |
| Big-5: Neuroticism | 2.92 | 2.94 | 2.86 | 0.08 | 0.372 | 442 |
| **D. Further control variables at baseline** | | | | | | |
| Big-5: Openness | 3.46 | 3.46 | 3.46 | 0.00 | 0.992 | 442 |
| Big-5: Extraversion | 3.35 | 3.33 | 3.40 | -0.07 | 0.460 | 441 |
| Big-5: Agreeableness | 3.47 | 3.48 | 3.43 | 0.05 | 0.532 | 441 |
| Higher SES | 0.54 | 0.54 | 0.53 | 0.01 | 0.811 | 442 |

Notes: Table shows group means for the total (column 1), randomized (column 2), and non-randomized (column 3) samples in the baseline survey. Column 4: difference between the averages of the randomized and non-randomized sample. Column 5: *p*-value of the coefficient on the randomized-sample indicator in a regression that regresses the specific variable on the randomized-sample indicator.

## Appendix C: Pair-wise Randomization Design

This appendix describes our pair-wise randomization approach. To achieve randomization of participants into treatment and control groups, we implemented a pair-wise matching design followed by rerandomization within the matched pairs, using the computationally feasible optimal greedy algorithm (Bruhn and McKenzie (2009)). Pair-wise matching designs with rerandomization have desirable statistical properties compared to a simple unconditional single-draw randomization procedure (e.g., Greevy et al. (2004); Imai, King, and Nall (2009); Bruhn and McKenzie (2009); Morgan and Rubin (2012); Kasy (2016); Imbens and Rubin (2015)). In particular, they achieve higher statistical power, avoid substantial imbalance in observable characteristics by chance in small samples, and improve the possibilities to investigate the robustness of results in case of attrition in later survey waves (see Appendix D).[3]

We conducted the randomization separately for each site. This was steered by the fact that the official starting date of the program varied slightly across sites. We fielded the baseline survey briefly before the site-specific program start and conducted the randomization during the few days between baseline survey and program start. The separate randomization for each site ensured perfect matching on regional and local circumstances.

The randomization process included three steps. The first step is the pair-wise matching. We matched statistical pairs of applicants by minimizing the (scale-invariant) Mahalanobis distance between the values of a vector of matching variables $\boldsymbol{X}$ between observations $i$ and $j$ within pairs:

$$\Delta(\boldsymbol{X}_i, \boldsymbol{X}_j) = \sqrt{(\boldsymbol{X}_i - \boldsymbol{X}_j)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{X}_i - \boldsymbol{X}_j)} \tag{C1}$$

where $\boldsymbol{\Sigma}^{-1}$ denotes the inverse of the covariance matrix.[4]

As the quality of the balancing for each variable deteriorates as more variables are included in the randomization process, we restricted the set of baseline variables considered in the

[3] If treatment effects are not homogenous and drop-out is related to the size of the treatment effect, dropping a pair unit yields a consistent estimate of the average treatment effect for the subsample of units that remain in the sample, not for the full sample (Bruhn and McKenzie (2009)).

[4] To implement our randomization, we adopted the Stata code provided in the supplementary material of Bruhn and McKenzie (2009).

matching to variables that are expected to both be highly predictive of future outcomes and have a low share of missing values.[5] The selected covariates for the pair-wise matching are gender, classroom, and baseline grades in math and German (coarsened from six to three distinct values).[6] In cases of uneven numbers of applications at a site, the size of the last matched group was increased to three in order to avoid a single remaining observation.

The second step is to generate a set of random treatment allocations. We ran 1,000 replications in which we randomly assigned one individual within each pair to the treatment group and the other to the control group. To evaluate the balancing after each rerandomization, we computed balancing statistics for the following eleven variables observed in the baseline survey: age, gender, migrant status, books at home (categories), self-reported grades in math, German, and English, an indicator for receiving paid private teaching, parental homework support, neuroticism, and conscientiousness.

For each replication $s$, we estimated bivariate regressions of each baseline variable $X_k$ on a treatment indicator $T$. To detect the presence of a statistically significant difference in a baseline variable between the treatment and control groups, we computed the $p$-value of the estimate $\beta_{ks}$ on the treatment indicator:

$$X_{ks} = \alpha + \beta_{ks} T_s + \epsilon_{ks} \tag{C2}$$

To obtain an estimate for the size of the difference in baseline variables (economic significance), we computed the standardized bias:

$$bias_{X_{ks}} = 100 \cdot \frac{\bar{X}_{ks;T=1} - \bar{X}_{ks;T=0}}{\sqrt{\frac{\sigma^2_{ks;T=1} + \sigma^2_{ks;T=0}}{2}}} \tag{C3}$$

where $\bar{X}_{ks;T=0,1}$ and $\sigma^2_{ks,T=0,1}$ denote the estimated mean and variance, respectively, for baseline variable $X_k$ in replication $s$ computed separately for the control ($T = 0$) and treatment ($T = 1$)

[5] Due to the expectation of missing values, we did not consider parental education reported by adolescent applicants in the matching. In cases of missing values in the selected matching variables, a missing dummy was included in the randomization process.

[6] In the pair-wise matching, we used self-reported grades from the baseline survey as we did not yet have administrative report-card information when implementing the randomization. Treatment assignment had to be achieved within at most two weeks to not delay the start of the program, whereas some schools needed several months to grant us access to the administrative data.

groups. A high *p*-value and low bias define good balancing of a baseline variable across control and treatment groups.

The third step is to select the best replication based on balancing criteria. We chose the iteration that provided the best balancing, where the quality of the balancing of a replication is defined by the size of the minimum of the *p*-values and the maximum of bias associated with a single variable within a replication. We selected the allocation with the highest *p*-value minimum. In the case of a tie, we selected the replication with the lowest bias maximum.

Because we did not want to reduce the number of available slots in the program, sites without full oversubscription (2x number of applicants > number of available slots) occurred frequently.[7] In these cases, we had to assign both observations of some statistical pairs to the treatment group and, therefore, we lost these observations for the identification of treatment effects. We started to treat both observations in pairs with the worst match quality (highest Mahalanobis distance) until the size of the treatment group coincided with the available slots.

As some sites only allowed same-sex mentoring relations, we adjusted the rerandomization procedure for those sites to achieve a determined gender composition in the treatment group. In practice, this restriction of the set of treatment allocations had only little influence as gender was also used to form the matched pairs, and the gender constraint only restricted the set of potential randomization outcomes within gender-mixed pairs. This site-specific gender restriction never led to a deterministic outcome of the rerandomization process. After the restriction of the set of treatment allocations, the remaining allocations were compared with respect to their balancing and the allocation with the best balance was chosen. Although the gender composition was simultaneously and independently determined by the gender composition among the adolescent applicants and the available mentors, and therefore as good as randomly determined, we control for gender in our main specifications.

[7] Because mentor and mentee are typically required to be of the same gender, we essentially had to rely on gender-specific oversubscription.

## Appendix D: Attrition Analysis

This appendix presents the extent of sample attrition in our data and investigates whether sample attrition is selective.

Table D1 shows the absolute and relative numbers of recontacted observations in the full randomized sample (panel A) and by SES status (panels B and C), in total and separately for treatment and control groups. For the follow-up data, recontact rates are shown separately for the questionnaire data ("follow-up surveys") and the administrative school grade data ("administrative data").

In general, attrition one year after program start is extremely low. In every subsample, we were able to achieve recontact rates of 89 percent or higher. In fact, in almost all cases, recontact rates are above 95 percent. We only observe a slightly smaller recontact rate for low-SES control-group individuals in the follow-up survey one year after program start (89 percent). Combining survey and administrative data ("combined one-year follow-up data"), as we do for our main labor-market prospect index, we always have at least some information for more than 98 percent of the sample one year after program start. In the follow-up surveys two and three years after program start, we still managed to keep attrition low. Specifically, in the follow-up survey three years after program start, which was carried out as an online questionnaire, we were able to recontact 88 percent of the adolescents. In no subsample did the recontact rate fall below 80 percent.

Results in Table D2 show that attrition one year after program start is not selective with respect to treatment status or observables in the baseline period in the full sample or in the subsamples of low-SES and higher-SES adolescents. The table regresses an attrition indicator on the treatment indicator, the higher-SES indicator, the index of labor-market prospects in the baseline period, and their interactions. Across samples, coefficients are typically close to zero and statistically insignificant. Similarly, results in Table D3 show that attrition in the follow-up surveys two and three years after program start is independent of treatment status and is not selective with respect to baseline outcomes.

One feature of the pair-wise randomization design (see Appendix C) is that we can exclude pairs where (at least) one individual could not be reached in the follow-up, which preserves internal validity (Bruhn and McKenzie (2009)). Table D4 shows the results of this exercise for the index of labor-market prospects. As expected given the low attrition and the inclusion of randomization-pair fixed effects in the baseline model, there are basically no differences between the results from the main specification in column 1 and the other model, in which we drop pairs with at least one attriter one year after program start in the sample combining survey and administrative data (column 2) as well as in the survey sample (column 3) and administrative sample (column 4). Reassuringly, effects are also very similar to the baseline model when we drop pairs with at least one attriter in the follow-up survey three years after program start (column 5), indicating non-selective sample attrition even by that time.

In sum, the attrition analysis confirms that attrition is very low in our study and that there is no selective attrition that would give rise to identification issues.

**Table D1: Sample Observations**

| | Total (1) | Treatment (2) | Control (3) |
|---|---|---|---|
| **A. Full randomized sample** | | | |
| Baseline survey | 308 *(100%)* | 153 *(100%)* | 155 *(100%)* |
| Combined one-year follow-up data[a] | 304 *(98.7%)* | 152 *(99.4%)* | 152 *(98.0%)* |
| Follow-up surveys | | | |
| 1 year after program start | 291 *(94.5%)* | 146 *(95.4%)* | 145 *(93.5%)* |
| 2 years after program start | 261 *(84.7%)* | 130 *(85.0%)* | 131 *(84.5%)* |
| 3 years after program start | 272 *(88.3%)* | 137 *(89.5%)* | 135 *(87.1%)* |
| Administrative data | | | |
| ¾ years after program start | 294 *(95.5%)* | 146 *(95.4%)* | 148 *(95.5%)* |
| 2 years after program start (final grades) | 292 *(94.8%)* | 146 *(95.4%)* | 146 *(94.2%)* |
| **B. Low-SES** | | | |
| Baseline survey | 141 *(100%)* | 66 *(100%)* | 75 *(100%)* |
| Combined one-year follow-up data[a] | 138 *(97.9%)* | 65 *(98.5%)* | 73 *(97.3%)* |
| Follow-up surveys | | | |
| 1 year after program start | 131 *(92.9%)* | 64 *(97.0%)* | 67 *(89.3%)* |
| 2 years after program start | 109 *(77.3%)* | 52 *(78.8%)* | 57 *(76.0%)* |
| 3 years after program start | 119 *(84.4%)* | 59 *(89.4%)* | 60 *(80.0%)* |
| Administrative data | | | |
| ¾ years after program start | 131 *(92.9%)* | 62 *(93.9%)* | 69 *(92.0%)* |
| 2 years after program start (final grades) | 131 *(92.9%)* | 62 *(93.9%)* | 69 *(92.0%)* |
| **C. Higher-SES** | | | |
| Baseline survey | 167 *(100%)* | 87 *(100%)* | 80 *(100%)* |
| Combined one-year follow-up data[a] | 166 *(99.4%)* | 87 *(100%)* | 79 *(98.8%)* |
| Follow-up surveys | | | |
| 1 year after program start | 160 *(95.8%)* | 82 *(94.3%)* | 78 *(97.5%)* |
| 2 years after program start | 152 *(91.0%)* | 78 *(89.7%)* | 74 *(92.5%)* |
| 3 years after program start | 153 *(91.6%)* | 78 *(89.7%)* | 75 *(93.8%)* |
| Administrative data | | | |
| ¾ years after program start | 163 *(97.6%)* | 84 *(96.6%)* | 79 *(98.8%)* |
| 2 years after program start (final grades) | 161 *(96.4%)* | 84 *(96.6%)* | 77 *(96.3%)* |

Notes: Table shows observation numbers and relative resurvey probabilities (in parentheses) by treatment status, data wave, and SES background. [a] Survey one year after program start and administrative data ¾ years after program start.

**Table D2: Attrition Analysis One Year after Program Start**

| | Combined one-year follow-up data (survey and administrative) | | | | Follow-up survey one year after program start | | | | Administrative data ¾ years after program start | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Treatment | -0.013 | -0.013 | -0.025 | -0.024 | -0.022 | -0.020 | -0.093** | -0.089* | 0.000 | 0.001 | -0.019 | -0.021 |
| | *(0.009)* | *(0.009)* | *(0.019)* | *(0.019)* | *(0.023)* | *(0.024)* | *(0.046)* | *(0.046)* | *(0.018)* | *(0.019)* | *(0.038)* | *(0.040)* |
| Treatment x Outcome index in $t_0$ | | 0.000 | | 0.002 | | 0.019 | | 0.034 | | 0.009 | | -0.010 |
| | | *(0.013)* | | *(0.023)* | | *(0.022)* | | *(0.038)* | | *(0.023)* | | *(0.037)* |
| Outcome index in $t_0$ | | -0.000 | | 0.001 | | 0.006 | | 0.011 | | -0.001 | | 0.022 |
| | | *(0.012)* | | *(0.023)* | | *(0.023)* | | *(0.032)* | | *(0.024)* | | *(0.035)* |
| Treatment x Higher-SES | | | 0.023 | 0.021 | | | 0.131** | 0.123* | | | 0.039 | 0.040 |
| | | | *(0.021)* | *(0.021)* | | | *(0.064)* | *(0.064)* | | | *(0.050)* | *(0.054)* |
| Higher-SES | | | -0.026 | -0.026 | | | -0.067 | -0.056 | | | -0.054 | -0.051 |
| | | | *(0.019)* | *(0.019)* | | | *(0.042)* | *(0.045)* | | | *(0.033)* | *(0.035)* |
| Treatment x Higher-SES x Outcome index in $t_0$ | | | | -0.009 | | | | -0.031 | | | | 0.035 |
| | | | | *(0.020)* | | | | *(0.048)* | | | | *(0.051)* |
| Higher-SES x outcome index in $t_0$ | | | | -0.001 | | | | -0.010 | | | | -0.046 |
| | | | | *(0.021)* | | | | *(0.043)* | | | | *(0.040)* |
| Randomization-pair fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Covariates | No | No | No | No | No | No | No | No | No | No | No | No |
| Observations | 308 | 308 | 308 | 308 | 308 | 308 | 308 | 308 | 308 | 308 | 308 | 308 |
| $R^2$ | 0.75 | 0.75 | 0.75 | 0.75 | 0.59 | 0.59 | 0.60 | 0.60 | 0.70 | 0.70 | 0.70 | 0.71 |
| Attrition *F*-test *p*-values | | | | | | | | | | | | |
| Overall sample | 0.151 | 0.361 | 0.356 | 0.730 | 0.347 | 0.487 | 0.117 | 0.344 | 1.000 | 0.922 | 0.658 | 0.928 |
| Low-SES sample | | | 0.197 | 0.315 | | | 0.046 | 0.345 | | | 0.615 | 0.540 |
| Higher-SES sample | | | 0.769 | 0.470 | | | 0.234 | 0.404 | | | 0.378 | 0.421 |

Notes: Table shows attrition analysis within randomization pairs one year after program start. Dependent variable in columns 1-4, 5-8, and 9-12 is a dummy indicating attrition in the overall, survey, and administrative sample, respectively. *F*-test *p*-values report *p*-values from joint significance tests of all treatment-related coefficients for the indicated sample. Robust standard errors in parentheses. Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1.

**Table D3: Attrition Analysis Two and Three Years after Program Start**

| | Follow-up survey two years after program start | | | | Follow-up survey three years after program start | | | | Administrative data two years after program start | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Treatment | -0.011 | -0.015 | -0.067 | -0.073 | -0.033 | -0.032 | -0.137* | -0.129* | -0.013 | -0.014 | -0.022 | -0.025 |
| | (0.041) | (0.041) | (0.079) | (0.078) | (0.040) | (0.040) | (0.076) | (0.077) | (0.022) | (0.022) | (0.039) | (0.040) |
| Treatment x Outcome index in $t_0$ | | -0.072 | | -0.096 | | 0.015 | | 0.059 | | 0.001 | | -0.006 |
| | | (0.049) | | (0.079) | | (0.049) | | (0.078) | | (0.026) | | (0.039) |
| Outcome index in $t_0$ | | 0.017 | | 0.025 | | -0.001 | | -0.042 | | -0.008 | | 0.011 |
| | | (0.040) | | (0.058) | | (0.033) | | (0.051) | | (0.026) | | (0.037) |
| Treatment x Higher-SES | | | 0.114 | 0.118 | | | 0.193* | 0.183* | | | 0.022 | 0.021 |
| | | | (0.105) | (0.106) | | | (0.103) | (0.104) | | | (0.059) | (0.059) |
| Higher-SES | | | -0.147* | -0.167** | | | -0.107 | -0.103 | | | -0.061 | -0.059 |
| | | | (0.078) | (0.080) | | | (0.075) | (0.075) | | | (0.041) | (0.043) |
| Treatment x Higher-SES x Outcome index in $t_0$ | | | | 0.022 | | | | -0.103 | | | | 0.001 |
| | | | | (0.111) | | | | (0.109) | | | | (0.050) |
| Higher-SES x outcome index in $t_0$ | | | | -0.011 | | | | 0.088 | | | | -0.033 |
| | | | | (0.075) | | | | (0.063) | | | | (0.048) |
| Randomization-pair fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Covariates | No | No | No | No | No | No | No | No | No | No | No | No |
| Observations | 308 | 308 | 308 | 308 | 308 | 308 | 308 | 308 | 308 | 308 | 308 | 308 |
| $R^2$ | 0.50 | 0.51 | 0.51 | 0.52 | 0.42 | 0.42 | 0.44 | 0.44 | 0.61 | 0.61 | 0.61 | 0.61 |
| Attrition *F*-test *p*-values | | | | | | | | | | | | |
| Overall sample | 0.790 | 0.317 | 0.549 | 0.243 | 0.408 | 0.712 | 0.165 | 0.376 | 0.559 | 0.816 | 0.838 | 0.949 |
| Low-SES sample | | | 0.399 | 0.079 | | | 0.075 | 0.481 | | | 0.568 | 0.600 |
| Higher-SES sample | | | 0.385 | 0.770 | | | 0.287 | 0.915 | | | 0.994 | 0.763 |

Notes: Table shows attrition analysis within randomization pairs two and three years after program start. Dependent variable is a dummy indicating attrition in the data indicated in the respective column header. *F*-test *p*-values report *p*-values from joint significance tests of all treatment-related coefficients for the indicated sample. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

**Table D4: Effect of the Mentoring Program: Dropping Attrition Pairs**

| | Baseline | Dropping pairs with at least one attriter in | | | |
|---|---|---|---|---|---|
| | | Combined one-year follow-up data | Follow-up survey one year after program start | Admin. data ¾ years after program start | Follow-up survey three years after program start |
| | (1) | (2) | (3) | (4) | (5) |
| Treatment | 0.644*** | 0.644*** | 0.683*** | 0.650*** | 0.640*** |
| | *(0.149)* | *(0.148)* | *(0.157)* | *(0.155)* | *(0.171)* |
| Treatment x Higher-SES | -0.865*** | -0.865*** | -0.907*** | -0.916*** | -0.754*** |
| | *(0.227)* | *(0.226)* | *(0.240)* | *(0.238)* | *(0.262)* |
| Higher-SES | 0.286 | 0.286 | 0.273 | 0.271 | 0.166 |
| | *(0.192)* | *(0.192)* | *(0.203)* | *(0.197)* | *(0.203)* |
| Outcome in $t_0$ | 0.446*** | 0.446*** | 0.434*** | 0.445*** | 0.434*** |
| | *(0.076)* | *(0.076)* | *(0.081)* | *(0.079)* | *(0.087)* |
| Randomization-pair fixed effects | Yes | Yes | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes | Yes | Yes |
| Observations | 304 | 302 | 277 | 286 | 232 |
| $R^2$ | 0.477 | 0.479 | 0.463 | 0.483 | 0.496 |
| Treatment effect for Higher-SES | -0.221 | -0.221 | -0.225 | -0.266* | -0.113 |
| | *(0.136)* | *(0.136)* | *(0.143)* | *(0.142)* | *(0.152)* |

Notes: Table shows ITT effects of the mentoring program on the index of labor-market prospects when we drop pairs with at least one attriter one year after program start (columns 2-4) and three years after program start (column 5). Column 1 shows the baseline results from column 4 of Table 2. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in $t_0$ are included. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

## Appendix E: Labor-Market Analysis of Linked PIAAC and PIAAC-L Data

This appendix provides evidence from the German PIAAC/PIAAC-L dataset (Appendix E.1) on the association of school grades with cognitive skills and labor-market success in adulthood (Appendix E.2), the association of patience and trust with labor-market outcomes (Appendix E.3), and differences in professional qualifications by SES background (Appendix E.4).

### E.1 The PIAAC and PIAAC-L Data

The analyses of this appendix use the German sample of the Programme for the International Assessment of Adult Competencies (PIAAC) survey, a large-scale study administered by the OECD in 2011/2012 (OECD (2016)). In each participating country, a representative sample of at least 5,000 adults aged 16 to 65 years participated in PIAAC. In addition to information on tested cognitive skills, PIAAC provides data from an extensive background questionnaire with detailed information on respondents' demographic characteristics, educational degrees, and labor-market outcomes.

PIAAC was designed to measure key cognitive and workplace skills needed for individuals to advance in their jobs and participate in society. The survey assessed cognitive skills in three domains: numeracy, literacy, and problem-solving in technology-rich environments. The domains refer to key information-processing competencies. Numeracy skills are defined as the ability to access, use, interpret, and communicate mathematical information and ideas in order to engage in and manage the mathematical demands of a range of situations in adult life. Literacy skills are defined as the ability to understand, evaluate, use and engage with written texts to participate in society, to achieve one's goals, and to develop one's knowledge and potential. The domain of problem-solving in technology-rich environments, typically referred to as "ICT skills," is defined as the ability to use digital technology, communication tools, and networks to acquire and evaluate information, communicate with others, and perform practical tasks.[8] In the

[8] Not all respondents participated in the ICT-skills assessment, because of a lack of any computer experience, failing a short initial ICT test, or opting out of the domain (see Falck, Heimisch-Roecker, and Wiederhold (2021) for details).

empirical analysis, test scores in each domain are standardized with a mean of zero and a standard deviation of one.

Germany conducted a follow-up study, PIAAC-L, in which respondents who participated in the original German PIAAC study in 2011/2012 were interviewed in three further waves (2014, 2015, and 2016).[9] For this analysis, we focus on the first wave of PIAAC-L, which elicited more detailed information from the participants regarding their educational history, personality traits, and family background. In particular, respondents reported the grades in mathematics, German, and the first foreign language (typically English) from their last report card in secondary school.

### E.2 School Grades and Later-Life Outcomes

Our first PIAAC analysis provides descriptive evidence that math grades at the end of secondary school are strongly related to cognitive skills and labor-market success in adulthood.

In Table E1, we show how school grades are related to important adult outcomes. Columns 1-6 consider cognitive skills in numeracy, literacy, and ICT. Columns 7-12 focus on labor-market outcomes, investigating unemployment (columns 7-8) as well as monthly and hourly wages (columns 9-12).[10] Following our main specification to evaluate the impact of the mentoring program, we interact grades with an indicator for higher-SES to investigate whether grade effects differ by SES background.[11] Regressions control for demographic characteristics (a quadratic polynomial in age and gender) as well as school-type fixed effects. The grade scale is reversed, such that better grades indicate more beneficial outcomes, and standardized with a mean of zero and a standard deviation of one (normalized to the distribution of the estimation sample in column 1 of Table E1). The odd columns in the table include only math grades, the

[9] For a detailed description of the study design and the technical implementation of PIAAC-L, see Zabal, Martin, and Rammstedt (2016).

[10] All outcome variables are taken from the original PIAAC study, measured in 2011/2012. For both monthly and hourly wages, we trim the bottom and top 1 percent of the wage distribution to limit the influence of outliers (see Hanushek at al. (2015)). Hourly wages do not include bonuses and are not available for self-employed.

[11] The SES indicator is constructed similarly as in the main analysis (see section 4.1). Information on books at home, parental education, and migration status comes from PIAAC 2012; information on single parenthood comes from PIAAC-L 2014. To generate an indicator of respondents having a single parent, we use survey questions asking respondents how many of their first 15 years of life they have spent with both parents or guardians. We classify respondents as having lived in a single-parent household at age 15 when they have not lived the entire first 15 years of their life with both parents/guardians (biological or non-biological). While this is a rather conservative approach, still only 17 percent of respondents are classified as having lived in a single-parent household at age 15.

primary cognitive outcome measure in our evaluation study, in addition to the controls. The even columns also include German and foreign-language grades.[12]

The results in Table E1 show a clear pattern: math grades at the end of secondary school are a significant predictor of cognitive skills and labor-market success later in life. As expected, math grades are more strongly correlated with numeracy skills than with literacy and ICT skills in adulthood, but estimates are sizeable for all three cognitive outcomes. For respondents with a low-SES background, an improvement in math grades by one standard deviation is related to an increase in adult numeracy skills by 16.6 percent of a standard deviation (column 1), in adult literacy skills by 12.5 percent of a standard deviation (column 3), and in adult ICT skills by 12.3 percent of a standard deviation (column 5). The relationship between math grades and labor-market success is also strong. When math grades increase by one standard deviation, the unemployment rate of respondents with a low-SES background decreases by 1.2 percentage points (26 percent of the full-sample mean and 22 percent of the mean in the low-SES sample) (column 7), while their monthly wages increase by 7.9 percent (column 9) and their hourly wages by 4.2 percent (column 11).[13] The interaction between outcomes and the higher-SES indicator is small and typically insignificant.

When we also include German and foreign-language grades at school (even columns of Table E1), the math-grade estimates are barely affected. Most strikingly, German grades and foreign-language grades are only weakly, if at all, related to cognitive skills and labor-market success in adulthood when math grades are also included. While the coefficients on German grades are small and insignificant across all outcomes, foreign-language grades are modestly related to cognitive skills, but play no role for labor-market outcomes. These results indicate that math grades at school are far more relevant in predicting human-capital formation and labor-market success later in life than German or foreign-language grades. This provides a strong

[12] Since grades are missing for some respondents, either because they could not remember the grade or they did not take the respective subject in the final year of secondary school, we impute missing grades with a constant. Thus, for each outcome, the specification with all grades is based on the same number of observations as the specification with math grades alone. To ensure that the imputed data are not driving our results, all regressions include an indicator for each grade with missing data that equals one for imputed values, and zero otherwise.

[13] The larger math coefficient on monthly wages compared to hourly wages suggests that the math grade also affects labor supply. Auxiliary regressions support this conjecture, as we find a positive relationship of math grades with the number of hours works and an indicator of working full-time, both at the intensive margin (i.e., for those who are employed) and the extensive margin (i.e., in the full sample).

argument for focusing on math grades as a proxy for cognitive skills in the experimental analysis of the mentoring program.

### E.3 Behavioral Traits and Labor-Market Outcomes

Next, we investigate how labor-market success is associated with patience and trust – two main behavioral outcome measures in our evaluation. This is enabled by the fact that the 2014 wave of PIAAC-L elicited several dimensions of respondents' personality traits – grit, trust, the Big-5 personality traits, internal and external locus of control, and risk preferences.

Unfortunately, PIAAC-L did not assess individuals' patience directly. However, the concept of grit is strongly related to patience, as it is defined as "perseverance and passion for long-term goals" (Duckworth et al. (2007)). In Table E2, we link the PIAAC-L measure of grit[14] to labor-market outcomes assessed in PIAAC.[15] As in Table E1, we consider unemployment as well as monthly and hourly wages. In the odd columns, grit (as well as its interaction with an indicator for higher-SES background) is included together with standard demographic controls. In the even columns, we add the other personality traits (i.e., trust, Big-5, internal and external locus of control, and risk preferences) as further controls.[16]

Across specifications, grit is strongly related to labor-market outcomes. For respondents with a low-SES background, a one-standard-deviation increase in grit is related to a decrease in unemployment by 3.1 percentage points (67 percent of the full-sample mean and 55 percent of the low-SES-sample mean) (column 1), an increase in monthly wages by 10.1 percent (column 3), and an increase in hourly wages by 4.9 percent (column 5). The positive grit effects on unemployment and hourly wages are somewhat stronger for individuals with low-SES background, while the SES interaction is significant only for unemployment. When adding the

[14] Grit is measured by the extent respondents agree to the following questions (grit scale by Duckworth et al. (2007) and Duckworth and Quinn (2009)): "I am a hard worker;" "I am self-disciplined;" "I can cope with setbacks;" "I finish whatever I begin;" "I have difficulty maintaining focus on projects or tasks that take more than a few months to complete" (reversed). The scale of answers ranged from 1 (= not at all) to 5 (= to a very large extent). Our measure of grit is the simple average of responses to the five items. In the empirical analysis, grit (as well as all other personality traits) are standardized with mean zero and standard deviation one.

[15] Results are qualitatively similar when we use labor-market outcomes elicited in PIAAC-L 2015, i.e., one year after the personality traits were measured. However, we prefer to use the outcomes from PIAAC 2011/2012 due to the larger sample size, as it also includes individuals who could not be resurveyed between PIAAC-L 2014 (when personality traits were measured) and 2015 (when labor-market outcomes were measured again).

[16] There are very few missing values (seven in total) for the personality traits. We impute these missing values with a constant such that the models with and without other personality traits as controls are based on the same number of observations. All regressions include an indicator for each personality trait with missing data that equals one for imputed values, and zero otherwise.

other personality traits in the even columns, the grit coefficients even tend to increase for the wage outcomes. Among the other personality traits, higher values of trust (see below) and external locus of control are consistently related to better labor-market outcomes, and so are lower values of extraversion and agreeableness. However, with the exception of trust, effect sizes are smaller than those for grit.

There is little prior evidence on the importance of trust (which is part of our measure of social skills) for individuals' labor-market outcomes. One noticeable exception is the work by Butler, Giuliano, and Guiso (2016), who find a non-linear relationship between trust and household income in the European Social Survey (ESS). In the ESS, trust is measured using the question: "Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?" with answer categories on a scale from zero to ten. The authors find that for trust levels between zero and seven, an increase in trust is associated with higher household income; for higher levels of trust, more trust is associated with a decrease in income. In our sample of applicants to the mentoring program, the average baseline level of trust is below the "critical" value of seven identified by Butler, Giuliano, and Guiso (2016), at 6.2 in the treatment group and 6.3 in the control group.[17] Aggregate evidence supports a positive relationship between trust and income at the country level (Knack and Keefer (1997); Algan and Cahuc (2010)).

In Table E3, we assess the relationship between trust and individual labor-market outcomes.[18] The table is constructed analogously to Table E2. Across specifications, trust is strongly related to labor-market outcomes. For respondents with a low-SES background, a one-standard-deviation increase in trust is related to a decrease in unemployment by 2.5 percentage points (54 percent of the full-sample mean and 45 percent of the low-SES-sample mean) (column 1), an increase in monthly wages by 15 percent (column 3), and an increase in hourly wages by 8.8 percent (column 5). The interaction of trust with the higher-SES indicator suggests that trust

[17] Note that we use a question very similar to ESS to elicit trust (see section 4.2 in the main text): "In general one can trust people." Participants answered on an 11-point scale where zero means "does not apply at all" and ten means "applies completely".

[18] In PIAAC-L, trust is measured by the extent to which respondents agree with the following statements: "In general, you can trust other people;" "Nowadays one can't rely on anyone" (reversed); and "If one is dealing with strangers, it is better to be careful not to trust them" (reversed). The answer scales range from one (fully agree) to four (fully disagree). After taking the mean of the three trust items, we standardize the resulting trust index with mean zero and standard deviation one.

effects do not differ significantly by SES background. When the other personality traits are included in the even columns, the trust coefficient decreases somewhat, but remains statistically significant. Similar to Butler, Giuliano, and Guiso (2016), we find some evidence for a hump-shaped relationship between trust and wages, albeit the quadratic term is significant only for monthly wages (not shown).

Overall, these results suggest that higher levels of grit and trust are positively associated with individual economic performance.

### E.4 Professional Qualifications by SES Background and School Grades

Finally, we investigate how obtained professional qualifications differ by SES background and whether an improvement in math grades increases the likelihood to enter the labor market with a qualification.

We start with evidence supporting the idea that successfully completing an apprenticeship is a desirable outcome for the target group of the mentoring program. The upper panel of Table E4 documents a substantial SES gap in the probability of failing to obtain any professional qualification, i.e., completing neither an apprenticeship nor a university degree. Focusing on those aged over 35 years (who are likely to have completed their final educational degree), 20 percent of individuals with low-SES background have no professional qualification, compared to only 4 percent in the group of higher-SES individuals. Results are very similar when considering individuals aged over 25, 30, or 40 years, suggesting strong persistence over the lifecycle and thus a policy focus on alleviating SES differences early in the professional career.[19]

The large SES gap in successfully obtaining a professional qualification is partly due to the fact that individuals with a low-SES background are more likely to drop out of an apprenticeship than their higher-SES counterparts. Focusing on persons older than 30 years, most of whom have finished their formal education, the probability of individuals with a low-SES background to ever have dropped out of apprenticeship training is 8.2 percent, which is twice as large as the probability for individuals with a higher-SES background (4.1 percent) (middle panel of Table E4).[20] From those individuals having experienced an apprenticeship dropout, 60 percent in the low-SES sample above age 30 have not obtained any professional qualification, compared to in

[19] Since PIAAC is cross-sectional in nature, the lifecycle SES differences may also reflect cohort effects.

[20] The information on previous dropout episodes is not available for individuals who were still enrolled in formal education at the time of the PIAAC interview.

the higher-SES sample. The share of individuals with low-SES background aged above 30 who have completed an apprenticeship is somewhat smaller than the corresponding share for individuals with a higher-SES background (66 vs. 69 percent).

At the same time, only 9.6 percent of individuals with a low-SES background aged above 30 have obtained a university degree – compared to 25.9 percent of individuals with higher-SES background (bottom panel of Table E4). Further taking into account that 2.6 percent of individuals in the low-SES group above age 30 have experienced a university dropout (of whom 73 percent have not obtained a university degree),[21] the evidence suggests that university education is not a viable option for the overwhelming majority of individuals with low-SES background.

This evidence has important implications for the qualification outcomes to be considered in the evaluation of the mentoring program. The mentoring program is not designed to address the (apparently substantial) barriers to enter university for disadvantaged youths, which likely include lacking educational aspirations of parents, peer effects being absent or even negative due to low-ability peers, low school quality, and others. Moreover, since the mentoring program is targeted towards adolescents from lower-track secondary schools, even successfully finishing these schools does not provide a university entrance qualification. Accordingly, only 1.7 percent of individuals with a low-SES background who obtained their highest school-leaving certificate from low-track (*Hauptschulen*) or intermediate-track (*Realschulen*) secondary schools have a completed university degree at an age above 30; for individuals with a higher-SES background, this share is still only 6.3 percent (see bottom panel of Table E5).[22] Thus, entering university is simply no option for the vast majority of low-SES participants in the mentoring program, at least not in the short run. In the context of our study, the question is rather whether the mentoring program can help disadvantaged youths to find an apprenticeship after school and to successfully complete it.

[21] 6.1 percent of respondents above age 30 with higher-SES background have experienced university dropout, of whom 67 percent have not obtained a university degree.

[22] For respondents who are not currently in the formal education system, PIAAC and PIAAC-L collect information only on the *highest* secondary school degree. Therefore, we cannot observe whether individuals attended a lower-track secondary school before finishing a higher school track. Our sample of individuals with lower-track secondary education as their highest secondary school degree thus likely contains less able graduates from lower-track schools, and thus the results in Table E5 likely underestimate the probability of completing university for the entire population of graduates from lower-track schools.

In the paper, we document a strong effect of the mentoring program on math grades for low-SES participants. The linked PIAAC and PIAAC-L data allow us to investigate whether better grades at the end of secondary school are associated with better qualification outcomes in adulthood. This complements the analysis in Appendix E.2 of grade effects on employment and wages. We keep only individuals above the age of 30 to ensure that most of them have finished their formal education. Table E6 is organized analogously to Table E1: While the odd columns include math grades as the only grade variable, the even columns further add German and foreign-language grades. Outcomes are indicators of not having obtained any qualification (columns 1 and 2), of having successfully completed an apprenticeship (columns 3 and 4), and of having quit one or more apprenticeships (columns 5 and 6).

Results in Table E6 show that better math grades decrease the probability of both not having obtained any qualification and having experienced an apprenticeship dropout at an age above 30. In terms of magnitude, a one-standard-deviation increase in math achievement reduces the probability of not having obtained a qualification by 2.6 percentage points for individuals with low-SES background (36 percent of the full-sample mean and 16 percent of the mean in the low-SES sample) (column 1). The probability of apprenticeship dropout is reduced by 1.2 percentage points (24 percent of the full-sample mean and 15 percent of the mean in the low-SES sample) (column 5). Better math grades tend to be negatively related to the probability of completing an apprenticeship, but coefficients are small and at most marginally significant (columns 3 and 4). There is no evidence for SES heterogeneity in grade effects. Furthermore, neither German grades nor foreign-language grades are systematically related to qualification outcomes, conditional on math grades.

As the mentoring program is targeted towards lower-track secondary schools, we also investigate whether better grades improve the career prospects of individuals who graduated from these schools. Table E7 restricts the sample to individuals aged above 30 who obtained their highest school degree from low-track (*Hauptschulen*) or intermediate-track (*Realschulen*) secondary schools. In this sample, math grade effects on qualification outcomes are even stronger than in the full sample. In particular, a one-standard-deviation-increase in math achievement decreases the probability of not having obtained any qualification for individuals with a low-SES background by 4 percentage points, which corresponds to 59 percent of the mean in the lower-track secondary school sample (33 percent of the low-SES mean) (column 1). Better math grades also increase the probability of finishing an apprenticeship, although the effect magnitude is rather small (column 3). The probability to drop out of an apprenticeship training decreases in math grades, by 1.6 percentage points (26 percent of the sample mean, 18 percent of the low-SES mean) for a one-standard-deviation increase in math achievement (column 5). There is no apparent effect heterogeneity by SES background. As was the case in the full sample, grades in German and foreign language are not significantly related to qualification outcomes in the sample of lower-track school graduates.

This evidence suggests that individuals with better math grades at secondary school are better able to manage the transition to the labor market. This translates to more favorable labor-market outcomes later in life. Although this evidence is purely descriptive, it does suggest that by improving math grades at school, the mentoring program may put disadvantaged youths on more favorable career tracks.

**Table E1: Math Grades at School and Later-Life Outcomes**

| | Numeracy skills | | Literacy skills | | ICT skills | | Unemployed | | Monthly wage | | Hourly wage | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Math grade | 0.166*** | 0.163*** | 0.125*** | 0.118*** | 0.123*** | 0.119*** | -0.012*** | -0.012*** | 0.079*** | 0.077*** | 0.042*** | 0.045*** |
| | *(0.013)* | *(0.013)* | *(0.013)* | *(0.013)* | *(0.015)* | *(0.016)* | *(0.004)* | *(0.004)* | *(0.016)* | *(0.017)* | *(0.010)* | *(0.010)* |
| Math grade x Higher-SES | 0.006** | 0.004 | 0.005* | 0.003 | 0.001 | 0.000 | 0.001 | 0.000 | 0.004 | 0.001 | 0.003** | 0.002 |
| | *(0.002)* | *(0.003)* | *(0.002)* | *(0.003)* | *(0.003)* | *(0.004)* | *(0.001)* | *(0.001)* | *(0.003)* | *(0.003)* | *(0.002)* | *(0.002)* |
| Higher-SES | 0.369*** | 0.321*** | 0.413*** | 0.373*** | 0.378*** | 0.373*** | -0.000 | -0.003 | 0.125*** | 0.113*** | 0.058** | 0.060** |
| | *(0.034)* | *(0.036)* | *(0.036)* | *(0.037)* | *(0.041)* | *(0.043)* | *(0.011)* | *(0.011)* | *(0.040)* | *(0.043)* | *(0.024)* | *(0.025)* |
| German grade | | -0.005 | | 0.011 | | -0.000 | | -0.004 | | 0.027 | | 0.004 |
| | | *(0.014)* | | *(0.015)* | | *(0.017)* | | *(0.005)* | | *(0.019)* | | *(0.012)* |
| German grade x Higher-SES | | 0.000 | | -0.000 | | 0.001 | | 0.001 | | 0.006** | | 0.003** |
| | | *(0.002)* | | *(0.002)* | | *(0.003)* | | *(0.001)* | | *(0.003)* | | *(0.001)* |
| Foreign-language grade | | 0.035** | | 0.034** | | 0.029 | | 0.003 | | -0.012 | | -0.014 |
| | | *(0.015)* | | *(0.016)* | | *(0.018)* | | *(0.005)* | | *(0.019)* | | *(0.012)* |
| Foreign-language grade x Higher-SES | | 0.002** | | 0.002** | | -0.000 | | 0.000 | | -0.000 | | -0.001 |
| | | *(0.001)* | | *(0.001)* | | *(0.001)* | | *(0.000)* | | *(0.001)* | | *(0.001)* |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| School-type fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Grade imputation dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 3,758 | 3,758 | 3,758 | 3,758 | 3,228 | 3,228 | 3,019 | 3,019 | 2,637 | 2,637 | 2,419 | 2,419 |
| $R^2$ (adjusted) | 0.401 | 0.408 | 0.399 | 0.404 | 0.338 | 0.339 | 0.030 | 0.034 | 0.267 | 0.269 | 0.352 | 0.354 |

Notes: Ordinary least squares estimates. Dependent variables: numeracy skills (columns 1 and 2), literacy skills (columns 3 and 4), ICT skills (columns 5 and 6), dummy for unemployment (columns 7 and 8), log monthly wages (columns 9 and 10), and log hourly wages (columns 11 and 12). Cognitive skills and grades are standardized to have mean zero and standard deviation one. See section 4.1 for the definition of *Higher-SES*. Sample: respondents aged 16-65 years in PIAAC who participated in PIAAC 2011/2012 and PIAAC-L 2014. All specifications control for a quadratic polynomial in age, gender, as well as for fixed effects for the type of secondary school that respondents attended and imputation dummies for school grades. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Data sources: PIAAC 2011/2012, PIAAC-L 2014.

**Table E2: Grit and Labor-Market Outcomes**

| | Unemployed | | Monthly wages | | Hourly wages | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Grit | -0.031*** | -0.026*** | 0.101*** | 0.125*** | 0.049*** | 0.069*** |
| | *(0.010)* | *(0.010)* | *(0.032)* | *(0.032)* | *(0.018)* | *(0.018)* |
| Grit x Higher-SES | 0.026** | 0.025** | 0.021 | 0.012 | -0.010 | -0.008 |
| | *(0.011)* | *(0.010)* | *(0.037)* | *(0.036)* | *(0.022)* | *(0.021)* |
| Higher-SES | -0.014 | -0.005 | 0.248*** | 0.176*** | 0.174*** | 0.109*** |
| | *(0.009)* | *(0.011)* | *(0.036)* | *(0.036)* | *(0.022)* | *(0.022)* |
| Openness | | 0.001 | | -0.019 | | 0.017 |
| | | *(0.004)* | | *(0.017)* | | *(0.011)* |
| Conscientiousness | | -0.009* | | -0.016 | | -0.046*** |
| | | *(0.005)* | | *(0.019)* | | *(0.012)* |
| Extraversion | | 0.009** | | -0.041** | | -0.041*** |
| | | *(0.004)* | | *(0.017)* | | *(0.010)* |
| Agreeableness | | 0.012*** | | -0.076*** | | -0.046*** |
| | | *(0.004)* | | *(0.016)* | | *(0.010)* |
| Neuroticism | | 0.012** | | 0.004 | | -0.010 |
| | | *(0.005)* | | *(0.018)* | | *(0.011)* |
| Trust | | -0.013*** | | 0.125*** | | 0.093*** |
| | | *(0.004)* | | *(0.017)* | | *(0.010)* |
| Internal locus of control | | 0.004 | | -0.015 | | -0.006 |
| | | *(0.004)* | | *(0.017)* | | *(0.010)* |
| External locus of control | | -0.009** | | 0.082*** | | 0.070*** |
| | | *(0.005)* | | *(0.018)* | | *(0.011)* |
| Risk attitude | | 0.007 | | 0.008 | | -0.027*** |
| | | *(0.005)* | | *(0.017)* | | *(0.010)* |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes |
| Imputation dummies for personality traits | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 3,019 | 3,019 | 2,637 | 2,637 | 2,419 | 2,419 |
| $R^2$ (adjusted) | 0.021 | 0.037 | 0.246 | 0.283 | 0.280 | 0.341 |

Notes: Ordinary least squares estimates. Dependent variable is indicated in the column header. Sample: respondents aged 16-65 years in the PIAAC survey who participated in PIAAC 2011/2012 and PIAAC-L 2014. See section 4.1 for the definition of *Higher-SES*. All behavioral traits are standardized to have mean zero and standard deviation one. All specifications control for a quadratic polynomial in age, gender, as well as for imputation dummies for personality traits. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Data sources: PIAAC 2011/2012, PIAAC-L 2014.

**Table E3: Trust and Labor-Market Outcomes**

| | Unemployed | | Monthly wages | | Hourly wages | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Trust | -0.025*** | -0.022** | 0.150*** | 0.139*** | 0.088*** | 0.079*** |
| | *(0.009)* | *(0.009)* | *(0.032)* | *(0.032)* | *(0.020)* | *(0.019)* |
| Trust x Higher-SES | 0.014 | 0.012 | -0.026 | -0.019 | 0.021 | 0.019 |
| | *(0.010)* | *(0.010)* | *(0.037)* | *(0.037)* | *(0.023)* | *(0.022)* |
| Higher-SES | -0.008 | -0.002 | 0.211*** | 0.174*** | 0.146*** | 0.111*** |
| | *(0.009)* | *(0.009)* | *(0.036)* | *(0.036)* | *(0.022)* | *(0.022)* |
| Openness | | 0.001 | | -0.019 | | 0.017 |
| | | *(0.005)* | | *(0.017)* | | *(0.011)* |
| Conscientiousness | | -0.008 | | -0.016 | | -0.046*** |
| | | *(0.005)* | | *(0.019)* | | *(0.012)* |
| Extraversion | | 0.008** | | -0.041*** | | -0.041*** |
| | | *(0.004)* | | *(0.017)* | | *(0.010)* |
| Agreeableness | | 0.011*** | | -0.075*** | | -0.046*** |
| | | *(0.004)* | | *(0.016)* | | *(0.009)* |
| Neuroticism | | 0.013*** | | 0.004 | | -0.010 |
| | | *(0.005)* | | *(0.018)* | | *(0.011)* |
| Grit | | -0.007 | | 0.134*** | | 0.064*** |
| | | *(0.005)* | | *(0.020)* | | *(0.012)* |
| Internal locus of control | | 0.004 | | -0.015 | | -0.006 |
| | | *(0.004)* | | *(0.017)* | | *(0.010)* |
| External locus of control | | -0.009* | | 0.082*** | | 0.070*** |
| | | *(0.005)* | | *(0.018)* | | *(0.011)* |
| Risk attitude | | 0.007 | | 0.008 | | -0.027*** |
| | | *(0.005)* | | *(0.017)* | | *(0.010)* |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes |
| Imputation dummies for personality traits | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 3,019 | 3,019 | 2,637 | 2,637 | 2,419 | 2,419 |
| $R^2$ (adjusted) | 0.021 | 0.035 | 0.252 | 0.283 | 0.307 | 0.341 |

Notes: Ordinary least squares estimates. Dependent variable is indicated in the column header. Sample: respondents aged 16-65 years in the PIAAC survey who participated in PIAAC 2011/2012 and PIAAC-L 2014. See section 4.1 for the definition of *Higher-SES*. All behavioral traits are standardized to have mean zero and standard deviation one. All specifications control for a quadratic polynomial in age, gender, as well as for imputation dummies for personality traits. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Data sources: PIAAC 2011/2012, PIAAC-L 2014.

**Table E4: Professional Qualifications by SES Background**

| | Low-SES | Higher-SES | Difference | Observations | |
|---|---|---|---|---|---|
| | Mean | Mean | *p*-value | Low-SES | Higher-SES |
| | (1) | (2) | (3) | (4) | (5) |
| **No qualification** | | | | | |
| Above age 25 | 0.21 | 0.05 | *0.000* | 1,282 | 2,920 |
| Above age 30 | 0.21 | 0.04 | *0.000* | 1,151 | 2,528 |
| Above age 35 | 0.20 | 0.04 | *0.000* | 1,016 | 2,206 |
| Above age 40 | 0.18 | 0.04 | *0.000* | 856 | 1,874 |
| **Apprenticeship (age > 30)** | | | | | |
| Dropout | 0.08 | 0.04 | *0.000* | 1,127 | 2,475 |
| Successful completion | 0.66 | 0.69 | *0.068* | 1,151 | 2,528 |
| **University (age > 30)** | | | | | |
| Dropout | 0.03 | 0.06 | *0.000* | 1,127 | 2,475 |
| Successful completion | 0.10 | 0.26 | *0.000* | 1,151 | 2,528 |

Notes: Table shows group means by SES background. See section 4.1 for the definition of *Higher-SES*. Column 3 shows the *p*-value from a *t*-test comparing the mean of the respective variable across groups. Information on dropout is not available for individuals who were still enrolled in formal education at the time of the PIAAC interview. Statistics weighted by sampling weights. Data source: PIAAC 2011/2012.

**Table E5: Professional Qualifications by SES Background: Individuals with Lower-Track Secondary Education**

| | Low-SES | Higher-SES | Difference | Observations | |
|---|---|---|---|---|---|
| | Mean | Mean | *p*-value | Low-SES | Higher-SES |
| | (1) | (2) | (3) | (4) | (5) |
| **No qualification** | | | | | |
| Above age 25 | 0.13 | 0.05 | *0.000* | 491 | 1,220 |
| Above age 30 | 0.13 | 0.05 | *0.000* | 445 | 1,118 |
| Above age 35 | 0.12 | 0.05 | *0.001* | 405 | 1,033 |
| Above age 40 | 0.11 | 0.04 | *0.008* | 351 | 898 |
| **Apprenticeship (> age 30)** | | | | | |
| Dropout | 0.08 | 0.05 | *0.033* | 442 | 1,103 |
| Successful completion | 0.85 | 0.88 | *0.167* | 445 | 1,118 |
| **University (> age 30)** | | | | | |
| Dropout | 0.01 | 0.01 | *0.155* | 442 | 1,103 |
| Successful completion | 0.02 | 0.06 | *0.000* | 445 | 1,118 |

Notes: Table shows group means by SES background. Sample includes only individuals who obtained their highest school-leaving certificate from low-track (*Hauptschulen*) or intermediate-track (*Realschulen*) secondary schools. See section 4.1 for the definition of *Higher-SES*. Column 3 shows the *p*-value from a *t*-test comparing the mean of the respective variable across groups. Information on dropout is not available for individuals who were enrolled in formal education at the time of the PIAAC interview. Statistics weighted by sampling weights. Data source: PIAAC 2011/2012.

**Table E6: Math Grades at School and Professional Qualifications**

| | No qualification | | Apprenticeship completion | | Apprenticeship dropout | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Math grade | -0.026*** | -0.022*** | -0.016* | -0.012 | -0.012** | -0.010** |
| | *(0.005)* | *(0.005)* | *(0.009)* | *(0.009)* | *(0.005)* | *(0.005)* |
| Math grade x Higher-SES | -0.002 | -0.001 | 0.001 | 0.001 | 0.000 | -0.001 |
| | *(0.001)* | *(0.001)* | *(0.001)* | *(0.002)* | *(0.001)* | *(0.001)* |
| Higher-SES | -0.052*** | -0.029** | 0.027 | 0.009 | -0.022 | -0.034** |
| | *(0.014)* | *(0.014)* | *(0.019)* | *(0.020)* | *(0.013)* | *(0.015)* |
| German grade | | -0.010 | | 0.002 | | -0.006 |
| | | *(0.006)* | | *(0.009)* | | *(0.006)* |
| German grade x Higher-SES | | 0.000 | | -0.002* | | 0.001 |
| | | *(0.001)* | | *(0.001)* | | *(0.000)* |
| Foreign-language grade | | -0.006 | | -0.016 | | 0.000 |
| | | *(0.007)* | | *(0.010)* | | *(0.006)* |
| Foreign-language grade x Higher-SES | | -0.001* | | 0.001** | | 0.001** |
| | | *(0.000)* | | *(0.001)* | | *(0.000)* |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes |
| School-type fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Grade imputation dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 2,633 | 2,633 | 2,633 | 2,633 | 2,582 | 2,582 |
| $R^2$ (adjusted) | 0.159 | 0.174 | 0.322 | 0.330 | 0.041 | 0.047 |

Notes: Ordinary least squares estimates. Dependent variables: dummy for no professional qualification obtained (columns 1 and 2), dummy for completed apprenticeship (columns 3 and 4), and dummy for apprenticeship dropout. Grades are standardized to have mean zero and standard deviation one. See section 4.1 for the definition of *Higher-SES*. Sample: respondents aged 31-65 years in PIAAC who participated in PIAAC 2011/2012 and PIAAC-L 2014. Information on dropout is not available for individuals who were enrolled in formal education at the time of the PIAAC interview. All specifications control for a quadratic polynomial in age, gender, as well as for fixed effects for the type of secondary school that respondents attended and imputation dummies for school grades. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Data sources: PIAAC 2011/2012, PIAAC-L 2014.

**Table E7: Math Grades at School and Professional Qualifications: Individuals with Lower-Track Secondary Education**

| | No qualification | | Apprenticeship completion | | Apprenticeship dropout | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Math grade | -0.040*** | -0.033*** | 0.020* | 0.020* | -0.016** | -0.014* |
| | *(0.007)* | *(0.008)* | *(0.010)* | *(0.011)* | *(0.008)* | *(0.008)* |
| Math grade x Higher-SES | -0.001 | -0.003 | 0.000 | 0.004** | 0.000** | -0.000 |
| | *(0.001)* | *(0.002)* | *(0.002)* | *(0.002)* | *(0.000)* | *(0.001)* |
| Higher-SES | -0.045*** | -0.032* | 0.010 | -0.004 | -0.028* | -0.042** |
| | *(0.016)* | *(0.016)* | *(0.019)* | *(0.020)* | *(0.015)* | *(0.018)* |
| German grade | | -0.013* | | 0.003 | | -0.005 |
| | | *(0.008)* | | *(0.011)* | | *(0.009)* |
| German grade x Higher-SES | | 0.003 | | -0.005*** | | 0.000 |
| | | *(0.002)* | | *(0.002)* | | *(0.001)* |
| Foreign-language grade | | -0.005 | | -0.017 | | 0.001 |
| | | *(0.009)* | | *(0.012)* | | *(0.008)* |
| Foreign-language grade x Higher-SES | | -0.000 | | 0.001* | | 0.001** |
| | | *(0.000)* | | *(0.001)* | | *(0.000)* |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes |
| School-type fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Grade imputation dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 1,563 | 1,563 | 1,563 | 1,563 | 1,545 | 1,545 |
| $R^2$ (adjusted) | 0.092 | 0.101 | 0.011 | 0.018 | 0.038 | 0.042 |

Notes: Ordinary least squares estimates. Dependent variables: dummy for no professional qualification obtained (columns 1 and 2), dummy for completed apprenticeship (columns 3 and 4), and dummy for apprenticeship dropout. Grades are standardized to have mean zero and standard deviation one. See section 4.1 for the definition of *Higher-SES.* Sample: respondents aged 31-65 years in PIAAC who participated in PIAAC 2011/2012 and PIAAC-L 2014 and who obtained their highest school-leaving certificate from low-track (*Hauptschulen*) or intermediate-track (*Realschulen*) secondary schools. Information on dropout is not available for individuals who were enrolled in formal education at the time of the PIAAC interview. All specifications control for a quadratic polynomial in age, gender, as well as for fixed effects for the type of secondary school that respondents attended and imputation dummies for school grades. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Data sources: PIAAC 2011/2012, PIAAC-L 2014.

## Appendix F: Mediation Analysis

This appendix presents details of the implementation of the mediation analysis for the low-SES (Appendix F.1) and higher-SES samples (Appendix F.2).

### F.1 Mediation Analysis for the Low-SES Sample

The primary aim of the mediation analysis is to investigate mediating factors of the significant positive treatment effect for low-SES adolescents. As potential mediators, we choose three variables, each elicited in both the treatment and control group, that are related to facilitating low-SES adolescents' transition into professional life. By referring to schools, future orientation, and occupational orientation, the three mediators each relate to one of three components of our index of labor-market prospects.

The first variable, *Perceive school as useful for job*, measures whether the respondent agrees with the statement that things learned in school could be useful for future jobs. It is measured on a 4-point scale and is standardized with a control-group mean of zero and a control-group standard deviation of one. The second variable, *Talk with mentor about future*, is a dummy variable that takes a value of one if the respondent talks to a mentor or coach about the future, and zero otherwise. The third variable, *Mentor important for job choice*, is a dummy variable that takes a value of one if the respondent refers to information from a mentor or coach as being important for job choice, and zero otherwise. Respondents could answer the survey question on the importance of the mentor or coach as a source of information for job choice on a 4-point scale from "very unimportant" to "very important," which we aggregate into a dummy variable taking a value of one if the mentor or coach is regarded as "rather important" or "very important", and zero otherwise. The aggregation allows us to include the sizeable fraction of individuals (22 percent) who respond "I don't know" in the non-important category.

The left panel of Table F1 shows that the three mediator variables are significantly affected by the treatment in the low-SES subsample (equation 6).[23,24]

[23] To be able to use the full sample in the mediation analysis, missing mediator values are imputed by the average in the treatment and control group, respectively, in Tables G1 and G2.

[24] Unconditionally, 43 percent of treated adolescents and 5 percent of control adolescents on average mention a mentor or coach as an attachment figure for talking about their future. On average, 59 percent of treated adolescents and 29 percent of control adolescents mention a mentor or coach as someone whom they refer to in order to receive occupational information.

The first column of Table F2 shows the overall treatment effect on the index of labor-market prospects for the low-SES sample (equation 4). Columns 2-5 show results when adding the mediator variables first individually and then jointly (equation 5). In our empirical implementation, we combine the estimates of the different equations and calculate the explained and unexplained shares of the treatment effect by using the *nlcom* command in Stata.

Comparing the treatment coefficients in the models with the individual mediators (columns 2-4) to the baseline model (column 1) yields the shares attributed to the three mediators that are depicted in the upper three bars of Panel A of Figure 4. When considered individually, *Perceive school as useful for job* accounts for 7 percent of the overall treatment effect for low-SES adolescents, *Talk with mentor about future* accounts for 31 percent, and *Mentor important for job choice* accounts for 16 percent.

The model that includes all three mediators jointly accounts for 37 percent of the overall treatment effect on the index of labor-market prospects for low-SES adolescents (comparison of columns 1 and 5: 1-0.416/0.658 = 0.37). Using equation 7 to assign shares to the individual mediators in the joint specification (shown in the fourth bar of Panel A of Figure 4), it becomes obvious that adolescents having a mentor as an attachment figure to talk about their future is by far the most relevant among the three mediators considered. In fact, the effect of the mentor being important for job choice materializes almost completely through talking with the mentor about the future.

Panel B of Figure 4 shows the mediation analysis for each of the three components of the index of labor-market prospects, based on columns 6-11 of Table F2. In the analysis, we follow Heckman, Pinto, and Savelyev (2013) and Oreopoulos, Brown, and Lavecchia (2017) in dropping mediators that would have a negative contribution in explaining the treatment effect, as the relative importance of the other mediators would be overestimated otherwise. The three mediators account for between 29 and 60 percent of the treatment effects on the three individual components. Surprisingly, the treatment effect of the mentoring program on math achievement in school is not mediated through perceiving school as useful for jobs, but rather through talking with the mentor about the future and conceiving the mentor important for job choice. As expected, the treatment effect on patience and social skills is primarily mediated through talking with the mentor about the future, and the treatment effect on labor-market orientation is primarily mediated through conceiving the mentor important for job choice.

## F.2 Mediation Analysis for the Higher-SES Sample

Overall, the mentoring program has a *negative* impact on higher-SES adolescents. While the treatment effect is statistically insignificant in the higher-SES sample, a mediation analysis can still provide some indication of where any negative effect might stem from.

From the mediators considered in the low-SES sample, *Talk with mentor about future* and *Mentor important for job choice* are also significantly positively affected by the treatment in the higher-SES sample (results not shown). However, among higher-SES adolescents, these mediators are not significantly associated with labor-market prospects, and they certainly cannot explain the *negative* treatment effect on labor-market prospects in the higher-SES sample.

Instead, we consider two mediator variables that capture potential crowding-out of other potentially performance-enhancing activities for higher-SES adolescents. The first variable, *Activities in school*, is an average of the following school activities (each of which is measured by a dummy variable that takes a value of one if the adolescent is engaged in the respective activity, and zero otherwise): acting as class representative, working as peer mediator, acting as school representative, working for the school magazine, volunteering as school nurse, participating in the school music ensemble, participating in the school theater group, and participating in other school activity. The variable is standardized with a control-group mean of zero and a control-group standard deviation of one. The second variable, *Good grades are important*, measures the extent to which adolescents consider good grades in school as important. The variable is measured on a 5-point scale and is standardized with a control-group mean of zero and a control-group standard deviation of one.

Columns 4 and 5 of Table F1 show that for higher-SES adolescents, these two mediators are negatively affected by the treatment. Comparing columns 12 and 13 of Table F2 indicates that the two mediators can account for 60 percent (=1-(-0.047/-0.117)) of the (insignificant) negative treatment effect in the higher-SES sample. The decomposition analysis of equation 7 indicates that in the joint specification, the crowding-out of activities in school (45 percent) turns out to be by far the most relevant channel, while reduced consideration of the importance of good school grades (15 percent) contributes less to the treatment effect.

**Table F1: Effect of the Mentoring Program on Mediator Variables**

| | Low-SES sample | | | Higher-SES sample | |
|---|---|---|---|---|---|
| | Perceive school as useful for job | Talk with mentor about future | Mentor important for job choice | Activities in school | Good grades are important |
| | (1) | (2) | (3) | (4) | (5) |
| Treatment | 0.271$^{*}$ | 0.363$^{***}$ | 0.302$^{***}$ | -0.325$^{**}$ | -0.258$^{*}$ |
| | *(0.163)* | *(0.073)* | *(0.083)* | *(0.157)* | *(0.139)* |
| Outcome in $t_0$ | Yes | – | Yes | – | Yes |
| Randomization-pair fixed effects | No | No | No | No | No |
| Covariates | Yes | Yes | Yes | Yes | Yes |
| Observations | 141 | 141 | 141 | 167 | 167 |
| $R^2$ | 0.203 | 0.275 | 0.179 | 0.096 | 0.294 |

Notes: Table shows ITT effects of the mentoring program on mediator variables one year after program start for the samples of low-SES and higher-SES adolescents, respectively. Dependent variable in column 1 measures on a 4-point scale whether the respondent agrees with the statement that things learned in school could be useful for a job. Dependent variable in column 2 is one if the individual talks to the mentor about the future, and zero otherwise. Dependent variable in column 3 is one if the mentor is important or very important for receiving information for job choice, and zero otherwise. Dependent variable in column 4 is an average index of the following school activities (represented by a dummy variable that is one if true, and zero otherwise): acting as class representative, working as peer mediator, acting as school representative, working for the school magazine, volunteering as school nurse, participating in the school music ensemble, participating in the school theater group, and participating in other school activity. Dependent variable in column 5 measures on a 5-point scale to what extent the adolescent considers good grades in school as important. Missing outcome values are imputed by treatment-specific averages. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Columns 1, 3, and 5 additionally control for baseline values of the respective mediator variable. Dummies for missing values in $t_0$ are included. Robust standard errors in parentheses. Significance levels: $^{***}$ $p < 0.01$, $^{**}$ $p < 0.05$, $^{*}$ $p < 0.1$.

**Table F2: Effect of the Mentoring Program on Labor-Market Prospects Conditional on Mediator Variables**

| | Low-SES | | | | | | | | | | | Higher-SES | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Index of labor-market prospects | | | | | Math achievement | | Patience and social skills | | Labor-market orientation | | Labor-market prospects | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) |
| Treatment | 0.658*** | 0.610*** | 0.455** | 0.556*** | 0.416** | 0.201 | 0.082 | 0.486*** | 0.218 | 0.488*** | 0.348* | -0.117 | -0.047 |
| | *(0.169)* | *(0.172)* | *(0.180)* | *(0.182)* | *(0.184)* | *(0.130)* | *(0.134)* | *(0.168)* | *(0.174)* | *(0.179)* | *(0.183)* | *(0.128)* | *(0.124)* |
| Perceive school as useful for job | | 0.163** | | | 0.130* | | 0.011 | | 0.167* | | 0.048 | | |
| | | *(0.082)* | | | *(0.076)* | | *(0.080)* | | *(0.093)* | | *(0.085)* | | |
| Talk with mentor about future | | | 0.567*** | | 0.458** | | 0.220 | | 0.618*** | | | | |
| | | | *(0.191)* | | *(0.207)* | | *(0.177)* | | *(0.202)* | | | | |
| Mentor important for job choice | | | | 0.348** | 0.135 | | 0.123 | | | | 0.427*** | | |
| | | | | *(0.175)* | *(0.181)* | | *(0.163)* | | | | *(0.158)* | | |
| Activities in school | | | | | | | | | | | | | 0.164** |
| | | | | | | | | | | | | | *(0.063)* |
| Good grades are important | | | | | | | | | | | | | 0.071 |
| | | | | | | | | | | | | | *(0.068)* |
| Outcome in $t_0$ | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Randomization-pair fixed effects | No | No | No | No | No | No | No | No | No | No | No | No | No |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 138 | 138 | 138 | 138 | 138 | 131 | 131 | 131 | 131 | 131 | 131 | 166 | 166 |
| $R^2$ | 0.531 | 0.546 | 0.564 | 0.548 | 0.576 | 0.554 | 0.568 | 0.351 | 0.427 | 0.355 | 0.395 | 0.262 | 0.295 |

Notes: Table shows ITT effects of the mentoring program on labor-market prospects one year after program start controlling for mediator variables in the samples of low-SES and higher-SES adolescents, respectively. Dependent variable in columns 1-5 and 12-13 is the index of labor-market prospects. Dependent variable in columns 6-11 is the respective component indicated in the column header. See notes of Table F1 for variable definitions. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Mediators are excluded if they would have a negative contribution in explaining the treatment effect. Columns 1-11 additionally control for baseline values of *Mentor important for job choice* and *Perceive school as useful for job*. Columns 12 and 13 additionally control for baseline values of *Good grades are important*. Dummies for missing values in $t_0$ are included. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

## Appendix G: Cost-Benefit Analysis

This appendix provides a quantification of the benefits and costs of the mentoring program. We express benefits in terms of the expected gain in lifetime earnings from program participation. Since most participants have not yet entered the labor market, this analysis requires assumptions about how the estimated program effects on labor-market prospects translate into gains in actual earnings and how these gains evolve over the lifecycle.

***Present value of lifetime earnings.*** We use a representative dataset of German adults, PIAAC (see Appendix E), to calculate discounted lifetime earnings separately for low-SES and higher-SES adults. The SES measure is constructed similarly as in the main analysis, using information on books at home in adolescence, parental education, single-parent status in adolescence, and first-generation migrant status (see section 4.1 and Appendix E.2). We first calculate annual earnings by multiplying monthly earnings by 12, and express this value in 2017 Euros (the year of program start for most of the adolescents in our sample). We smoothen the actual earnings stream by using predicted earnings from a regression of earnings on a quartic polynomial in age. We take into account that the age of labor-market entry differs by qualification (with 18 being the earliest entry age)[25] and assign persons before hypothetical labor-market entry zero earnings. We assume that persons exit the labor market at age 65.[26] Finally, to correct for periods of unemployment, we assign unemployed in our sample the standard rate of unemployment benefits. We discount future earnings at a net annual rate of 1.5 percent, which is comprised of a gross discount rate of 3 percent (e.g., Chetty et al. (2011); Heckman, Pinto, and Savelyev (2013); Lavecchia, Oreopoulos, and Brown (2020)) and a rate of potential output growth of 1.5 percent (Hanushek and Woessmann (2011); Hanushek, Ruhose, and Woessmann (2017)).

Although we find positive program effects on several outcomes, the cost-benefit analysis relies on the math-grade estimates. We do so for three reasons. First, as discussed in section 4.2, math achievement at school is highly predictive of future earnings. Second, among the outcomes studied in the paper, we deem math achievement as a measure of cognitive skills as most

[25] We use the following mean labor-market-entry ages by highest qualification as observed in the German Microcensus (Piopiunik, Kugler, and Woessmann (2017)): no qualification: age 18; apprenticeship training: age 21; Bachelor's degree: age 24 (university of applied sciences) or age 25 (university); Master's degree or higher: age 26 (university of applied sciences) or age 27 (university).

[26] The legal retirement age in Germany varies between 65 and 67 years, depending on the year of birth.

important for future labor-market success. Third, we know of no representative dataset that contains all variables necessary to construct our main outcome measure, the index of labor-market prospects, and to assess the earnings benefits of an increase in this index. We use data on math grades from PIAAC, which elicits math grades from the end of formal schooling.

To take into account that the standard deviations in math grades differ between PIAAC and our mentee sample, the estimates of the program effect in the cost-benefit analysis use non-standardized math grades. Our baseline specification (see column 1 of Table 3) yields a significant program effect of 0.425 grade points for low-SES adolescents one year after program start (with no statistically significant program effect for higher-SES adolescents). Our persistence analysis in Figure 5 indicates that treatment effects on math grades fade out by about 40 percent but remain significant two years after program start, i.e., at the end of low-track secondary school (see columns 1 and 3 of Table A11). Taking this fade-out into account, we use an effect size of 0.255 (=0.6*0.425) for the benefit calculations in the low-SES sample.

In PIAAC, we find that an increase in math grades by one grade point is associated with a monthly wage increase of 7.4 percent for low-SES individuals in the baseline model (equivalent to column 9 of Appendix Table E1). Multiplying the present value of lifetime earnings by the gain in monthly wages through better math grades and by the fade-out-adjusted treatment effect on math grades, we estimate that the gain in discounted lifetime earnings from the program is about 13,500 EUR for low-SES participants.[27] Since program participation does not lead to significant grade effects for higher-SES adolescents, their earnings benefits are assumed to be zero. Weighting the benefits of low-SES and higher-SES participants by the sample share of the respective group, we arrive at an estimate of the overall earnings benefits of the program of about 6,200 EUR.

***Program costs.*** According to the program's annual report, its total organizational costs amounted to 1,046,750 EUR in 2017.[28] Our best estimate of the number of mentoring pairs in operation in 2017 is about 1,400.[29] Thus, direct program costs are roughly 750 EUR per mentee.

[27] Note that we take a static perspective by assuming that program participation leads to a one-time earnings gain over the lifecycle. Alternatively, we could allow that an increase in math grades puts participants on a higher earnings trajectory. Program benefits would likely be even larger in this dynamic perspective.

[28] See https://rockyourlife.de/transparenz/ (accessed September 28, 2020).

[29] While there is no exact data on the number of mentoring pairs in operation in 2017, official data indicate that 837 new mentoring pairs were initiated in 2017. In our data, two-thirds of the mentoring relationships are still active one year after formation, which leads us to an estimate of roughly 1,400 active pairs in 2017.

Mentors work for the program on a voluntary and unpaid basis. While the mentors' time thus does not generate any direct program costs, we can also quantify the opportunity costs of the voluntary work. The program management estimates that mentors spent a total of roughly 160,000 hours of voluntary work for the program. Assuming an hourly wage rate of 10.60 EUR (the wage rate of a Bachelor-student assistant at the University of Munich in 2017), the opportunity costs of the program are about 1,200 EUR per mentor.

***Benefit-cost ratios.*** Table G1 reports benefit-cost ratios for different assumptions regarding (a) the discount rate, (b) program costs (with or without opportunity costs), and (c) program participants (with or without higher-SES adolescents). In all cases, program benefits exceed the costs to a sizeable extent. In our preferred specification with a net discount rate of 1.5 percent and no opportunity costs, the estimated program benefits outweigh costs by as much as 18-to-1 (13,500 EUR/750 EUR) if the program was targeted only at low-SES adolescents. If the program would not preselect only low-SES adolescents, the benefit-to-cost ratio would be 8-to-1 (6,200 EUR/750 EUR). When opportunity costs are also considered, the program yields benefit-cost ratios of 7-to-1 and 3-to-1, respectively. The large differences in the benefit-cost ratios by target group of the program indicates that the program foregoes substantial gains by not properly pre-screening participants. In fact, benefit-cost ratios would more than double if the program would focus on the slightly less than half of its subject pool that can be considered most disadvantaged.

These calculations can obviously provide only rough benchmarks for the program benefits. On the one hand, the estimated benefit-cost ratios would be lower if we were to assume that the program in fact has negative effects for higher-SES adolescents. On the other hand, there are also several reasons for why the calculations may underestimate the full program benefits. First, we consider only program effects on math grades and ignore potential earnings gains that accrue from positive effects on other outcomes (e.g., patience and labor-market orientation). Second, we measure benefits only with respect to earnings and ignore other potential pecuniary and nonpecuniary benefits, such as improvements in well-being and health, that are more difficult to quantify. Third, we focus on benefits for the mentees alone, neglecting potential benefits arising for mentors. For instance, social volunteering may increase mentors' job prospects if potential employers regard it as a signal for social skills (Piopiunik et al. (2020)). Thus, we consider our estimates of benefit-cost ratios as lower bounds of the actual values.

**Table G1: Benefit-Cost Ratios**

| | Actual costs | | Actual costs and opportunity costs of voluntary work | |
|---|---|---|---|---|
| | Program targeted at low-SES | Untargeted program | Program targeted at low-SES | Untargeted program |
| Discount rate | (1) | (2) | (3) | (4) |
| 0.0% | 25.9-to-1 | 11.9-to-1 | 9.9-to-1 | 4.5-to-1 |
| 1.5% | 18.0-to-1 | 8.3-to-1 | 6.9-to-1 | 3.2-to-1 |
| 3.0% | 13.0-to-1 | 6.0-to-1 | 5.0-to-1 | 2.3-to-1 |

Notes: Table shows estimates of benefit-cost ratios for different discount rates and different assumptions regarding the costs of the program (without or with opportunity costs of voluntary work) and its target group (low-SES only or low-SES and higher-SES). Untargeted program estimates assume zero program effects for higher-SES participants.

## Appendix H: Scalability

In this appendix, we show the potential reach of the mentoring program and discuss the assumptions underlying our calculations (Appendix H.1). We also provide the calculations for funding the project by the city governments (Appendix H.2).

### H.1 Potential Reach of the Program

This appendix provides a calculation of the potential reach of the mentoring program by estimating the size of the target group in Germany that would likely profit from participating in the program (i.e., the potential demand). The following calculation uses data from the Federal Statistical Office of Germany ("Kommunale Bildungsdatenbank"), which provides the number of pupils for each city/county, year, and school track. Whenever possible, the data refer to the school term 2017/18, i.e., the time period of our intervention.[30] With these data, we proceed as follows (see Table H1 for detailed numbers by state):

1. Identify the pupil population in cities/counties with a university, i.e., cities/counties in which *Rock Your Life!* could potentially set up a mentoring site: university cities/counties comprise cities/counties with (public or private) universities or universities of applied sciences. The relevant pupil population are pupils who are in lower-track schools, i.e., comprehensive and community schools (*Gesamt- und Gemeinschaftsschule*) and secondary schools from the lowest two tracks (*Realschule* and *Hauptschule*).[31] These are the school tracks that *Rock Your Life!* usually offers their program to. Thus, we do not consider the highest academic track schools (*Gymnasium*) where *Rock Your Life!* is not present.
   → Number of pupils in university cities/counties in the lowest school tracks in the school year 2017/18: 1,709,013

[30] For some federal states, there are no data available for 2017/18, so we use the latest available data for these states: 2016/17 for North Rhine-Westphalia, 2015/16 for Rhineland-Palatinate, and 2010/11 for Mecklenburg-Western-Pomerania.

[31] In Berlin, the respective school type is called integrated secondary schools (*Integrierte Sekundarschulen*). In Brandenburg, Lower Saxony, Saxony and Bremen, it is called *Oberschule*. Besides, in Saxony-Anhalt and North Rhine-Westphalia, secondary schools (*Sekundarschulen*) is the relevant school type. In Hessen, relevant schools are in addition *Mittelstufenschulen* and *Integrierte Jahrgangsstufen* as well as *Förderstufen*. In Mecklenburg-Western-Pomerania and Schleswig-Holstein, they are called regional schools (*Regionale Schule*) and *Regelschule* in Thuringia.

Note that this number amounts to 69 percent of all pupils attending the lowest school tracks over all cities and counties in Germany (2,481,082).

2. Identify the number of pupils in cities/counties with a university and attending the lowest school tracks *per grade level* (e.g., number of pupils in grade 8): this number is not given in the data directly; thus, we have to make an assumption about the number of pupils attending each grade level. In general, *Realschulen* have six grade levels and *Hauptschulen* have five grade levels. The grade levels of *Gesamtschulen* and *Gemeinschaftsschulen* depend on the federal state and can have six to nine grade levels. Assuming that pupils are equally assigned across grade levels, we calculate for each federal state and school type the number of pupils per grade. The sum over these calculations gives an estimate of the number of pupils in grade 8, the main target grade of *Rock Your Life!*.
   → Number of pupils in university cities/counties in the lowest school tracks in the school year 2017/18 in 8th grade: 268,871

3. Identify the *disadvantaged* pupils in university cities/counties in the lowest school tracks in the school year 2017/18 in 8th grade: to do that, we assume that half of all grade 8 pupils are very disadvantaged and would likely profit from participating in *Rock Your Life!*. We use this share as about half the pupils in our sample can be classified as low-SES. Again, we do the calculation separately for each federal state and then take the sum over all estimates.
   → Number of disadvantaged pupils in university cities/counties in the lowest school tracks in the school year 2017/18 in 8th grade: **134,441**

This number is the potential reach of *Rock Your Life!* if the program is fully scaled and targeted to low-SES adolescents in low-track schools. This amounts to 21 percent of the entire 8th-grade cohort (comprising pupils from all school tracks) in Germany.

If we restrict the calculation only to already existing sites, the potential reach is equal to 36,383 adolescents. Compared to the 742 mentoring pairs that actually existed in 2017/18, *Rock Your Life!* covers only 2 percent (=742/36,383) of the potential mentees in cities with existing mentoring sites. Thus, there is substantial scope for scaling up the program.

However, there are two bottlenecks that *Rock Your Life!* might face when scaling up the program:

1. There is a limited supply of university students who can act as mentors.

   To get an idea of the potential number of mentors, we use data on incoming students in 2017 at all German universities ($N$ = 509,400). The data are again obtained from the Federal Statistical Office. Importantly, this calculation uses only the number of incoming students because *Rock Your Life!* usually introduces the program to the cohort of freshmen students during their first lectures. Moreover, it is usually the case that one mentor has only one mentoring relationship during her or his study time.

   Naturally, not all first-year students are willing to engage in voluntary work. According to the Federal Ministry of Family Affairs, Senior Citizens, Women, and Youth, 55.6 percent of all university students are engaged in some voluntary activities.[32] Among those students who are engaged in voluntary work, 26.4 percent are active in areas related to youth and social work.[33] Using this information to adjust the freshmen student pool, we arrive at the following number of potential mentors:

   0.556*0.264*509,400 = 74,772

   This number implies that the potential reach of *Rock Your Life!* (134,441 adolescents) exceeds the number of potential mentors. This result is in line with our observation of oversubscription at the evaluated sites. Thus, taking the supply of mentors into account, the reach of the program is equal to 12 percent of the entire 8$^{th}$-grade cohort in Germany.

[32] See Figure 6 in https://www.bmfsfj.de/resource/blob/119820/b06feba2db2c77e0bff4a24662b20c70/freiwilliges-engagement-junger-menschen-data.pdf [last retrieved 04/01/2022].

[33] See Figure 9 in https://www.bmfsfj.de/resource/blob/119820/b06feba2db2c77e0bff4a24662b20c70/freiwilliges-engagement-junger-menschen-data.pdf [last retrieved 04/01/2022]. The share of 26.4 percent who engage in voluntary work is calculated from the three areas: school and kindergarten *(Schule oder Kindergarten)* 8.5 percent; social area *(Sozialer Bereich)* 8.7 percent; and extracurricular youth work or educational work for adults *(Außerschul. Jugendarbeit oder Bildungsarbeit für Erwachsene)* 9.2 percent.

2. Not all adolescents who can potentially be reached by the program want to participate. The supply-side restriction may not be important when we consider the fact that adolescents participate voluntarily in the program, and many may decide not to do so. Thus, each mentee can have a mentor if 44 percent of adolescents (=1-74,772/134,441) in the potential reach of the program would not participate. Then, the number of established mentoring relationships each year is around **75,000**.

## Table H1: Calculation of the Potential Reach of *Rock Your Life!*

| | BW (1) | BY (2) | B (3) | BB (4) | HB (5) | HH (6) | H (7) | MW (8) | LS (9) | NRW (10) | RP (11) | S (12) | SX (13) | SA (14) | SH (15) | T (16) | **Sum (17)** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Federal state (overall)** | | | | | | | | | | | | | | | | | |
| Pupils at low-track schools | 327,214 | 365,453 | 80,423 | 54,880 | 22,460 | 56,339 | 164,057 | 39,589 | 268,581 | 633,924 | 123,653 | 30,216 | 96,663 | 57,398 | 98,569 | 61,663 | **2,481,082** |
| Pupils at low-track schools per cohort | 56,674 | 60,909 | 11,489 | 12,019 | 2,808 | 6,260 | 32,808 | 6,599 | 41,491 | 92,635 | 18,321 | 3,418 | 16,111 | 8,665 | 16,428 | 8,859 | **395,494** |
| 50% of pupils at low-track schools | 28,337 | 30,455 | 5,745 | 6,010 | 1,404 | 3,130 | 16,404 | 3,300 | 20,746 | 46,318 | 9,161 | 1,709 | 8,056 | 4,333 | 8,214 | 4,430 | **197,752** |
| **University cities** | | | | | | | | | | | | | | | | | |
| Pupils at low-track schools | 235,195 | 167,818 | 80,423 | 34,442 | 17,659 | 56,339 | 115,405 | 9,240 | 136,230 | 583,577 | 66,554 | 5,020 | 84,257 | 38,610 | 50,193 | 28,051 | **1,709,013** |
| Pupils at low-track schools per cohort | 39,575 | 27,977 | 11,489 | 7,264 | 2,207 | 6,260 | 23,372 | 1,119 | 20,789 | 86,480 | 9,798 | 589 | 14,045 | 5,751 | 8,365 | 3,791 | **268,871** |
| 50% of pupils at low-track schools | 19,788 | 13,989 | 5,745 | 3,632 | 1,104 | 3,130 | 11,686 | 560 | 10,395 | 43,240 | 4,899 | 295 | 7,023 | 2,876 | 4,183 | 1,896 | **134,441** |
| ***Rock Your Life!* (RYL) Sites** | | | | | | | | | | | | | | | | | |
| Pupils at low-track schools per cohort | 9,285 | 8,651 | 11,489 | 0 | 0 | 6,260 | 5,065 | 0 | 3,841 | 18,863 | 669 | 0 | 4,203 | 1,715 | 2,715 | 0 | **72,756** |
| 50% of pupils at low-track schools | 4,643 | 4,326 | 5,745 | 0 | 0 | 3,130 | 2,533 | 0 | 1,921 | 9,432 | 335 | 0 | 2,102 | 858 | 1,358 | 0 | **36,383** |
| Share of pupils that attend school in university cities | 70% | 46% | 100% | 60% | 79% | 100% | 71% | 17% | 50% | 93% | 53% | 17% | 87% | 66% | 51% | 43% | **68.0%** |
| Current participation in RYL (RYL sites) | 180 | 87 | 33 | 0 | 0 | 20 | 28 | 0 | 29 | 200 | 8 | 0 | 47 | 43 | 67 | 0 | **742** |
| Share of federal state's students participating in RYL | 4% | 2% | 1% | 0% | 0% | 1% | 1% | 0% | 2% | 2% | 2% | 0% | 2% | 5% | 5% | 0% | **2.0%** |

Notes: Table shows the calculation of the potential reach of *Rock Your Life!* for each federal state (BW= Baden-Wurttemberg, BY = Bavaria, B=Berlin, BB=Brandenburg, HB=Bremen, HH=Hamburg, H=Hesse, MW=Mecklenburg-Western-Pomerania (2010/11), NRW=North Rhine-Westphalia (2016/17), LS=Lower Saxony, RP=Rhineland Palatinate (2015/16), S=Saarland, SX=Saxony, SA=Saxony-Anhalt, SH=Schleswig-Holstein, T=Thuringia). Data comes from the Federal Statistical Office and refers to the school year 2017/18 if not stated otherwise.

### H.2 Funding Calculations

As examples of funding needed from local city governments to sustainably finance a fully-scaled mentoring program, we have performed the following to example calculations.

***Aachen.*** The calculation is based on Aachen's household budget plan for the year 2016, where the city of Aachen spent 77,423,800 EUR on child, youth, and family services (more explicitly, on helping the youth and their families and on child and youth work). Overall, 8,621 students began studying in Aachen in the winter term 2016/17, among which 1,265 (=0.556*0.264*8,621) could potentially engage in voluntary work (for details, see our calculations in the previous section). On the demand side, approximately 1,226 adolescents can potentially be reached by *Rock Your Life!* in the city region of Aachen (same assumptions as above). Thus, 1,226 mentee-mentor pairs could potentially have been formed in Aachen in 2016. The overall program costs (1,950 EUR * 1,226 = 2,390,700 EUR) make up about 3 percent of the household budget on helping the youth and their families and on child and youth work.

***Cologne.*** In Cologne, the household budget for children and youths was approximately 300 million EUR in 2016. In the winter term 2016/17, 12,953 students started studying at all types of universities in Cologne, which yields 1,901 potential mentors. Overall, 2,189 adolescents can be identified as having *Rock Your Life!* potential. Consequently, 1,901 mentoring pairs could have been formed in 2016, resulting in total costs of 3.71 million EUR, which is equivalent to 1.2 percent of the budget for children and youths.

## Appendix References


Algan, Yann, Pierre Cahuc (2010). Inherited Trust and Growth. *American Economic Review* 100 (5): 2060-2092.

Beierlein, Constanze, Anastassyia Kovaleva, Christoph J. Kemper, Beatrice Rammstedt (2012). Ein Messinstrument Zur Erfassung Subjektiver Kompetenzerwartungen: Allgemeine Selbstwirksamkeit Kurzskala (Asku). GESIS-Working Paper 2012|17. Mannheim: GESIS – Leibniz-Institut für Sozialwissenschaften.

Bruhn, Miriam, David McKenzie (2009). In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics* 1 (4): 200-232.

Butler, Jeffrey V., Paola Giuliano, Luigi Guiso (2016). The Right Amount of Trust. *Journal of the European Economic Association* 14 (5): 1155-1180.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, Danny Yagan (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. *Quarterly Journal of Economics* 126 (4): 1593-1660.

Duckworth, Angela L., Christopher Peterson, Michael D. Matthews, Dennis R. Kelly (2007). Grit: Perseverance and Passion for Long-Term Goals. *Journal of Personality and Social Psychology* 92 (6): 1087-1101.

Duckworth, Angela Lee, Patrick D. Quinn (2009). Development and Validation of the Short Grit Scale (Grit–S). *Journal of Personality Assessment* 91 (2): 166-174.

Falck, Oliver, Alexandra Heimisch-Roecker, Simon Wiederhold (2021). Returns to ICT Skills. *Research Policy* 50 (7): 104064.

Goodman, Robert (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry* 38 (5): 581-586.

Greevy, Robert, Bo Lu, Jeffrey H. Silber, Paul Rosenbaum (2004). Optimal Multivariate Matching before Randomization. *Biostatistics* 5 (2): 263-275.

Hanushek, Eric A., Jens Ruhose, Ludger Woessmann (2017). Economic Gains from Educational Reform by US States. *Journal of Human Capital* 11 (4): 447-486.

Hanushek, Eric A., Ludger Woessmann (2011). How Much Do Educational Outcomes Matter in OECD Countries? *Economic Policy* 26 (67): 427-491.

Heckman, James J. (2020). Randomization and Social Policy Evaluation Revisited. NBER Technical Working Paper 107 (revised). Cambridge, MA: National Bureau of Economic Research.

Heckman, James J., Rodrigo Pinto, Peter Savelyev (2013). Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review* 103 (6): 2052-2086.

Imai, Kosuke, Gary King, Clayton Nall (2009). The Essential Role of Pair Matching in Cluster-Randomized Experiments, with Application to the Mexican Universal Health Insurance Evaluation. *Statistical Science* 24 (1): 29-53.

Imbens, Guido W., Donald B. Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. New York, NY: Cambridge University Press.

Kasy, Maximilian (2016). Why Experimenters Might Not Always Want to Randomize, and What They Could Do Instead. *Political Analysis* 24 (3): 324-338.

Kling, Jeffrey R., Jeffrey B. Liebman, Lawrence F. Katz (2007). Experimental Analysis of Neighborhood Effects. *Econometrica* 75 (1): 83-119.

Knack, Stephen, Philip Keefer (1997). Does Social Capital Have an Economic Payoff? A Cross-Country Investigation. *Quarterly Journal of Economics* 112 (4): 1251-1288.

Lavecchia, Adam M., Philip Oreopoulos, Robert S. Brown (2020). Long-Run Effects from Comprehensive Student Support: Evidence from Pathways to Education. *American Economic Review: Insights* 2 (2): 209-24.

List, John A., Azeem M. Shaikh, Yang Xu (2019). Multiple Hypothesis Testing in Experimental Economics. *Experimental Economics* 22 (4): 773-793.

Morgan, Kari Lock, Donald B. Rubin (2012). Rerandomization to Improve Covariate Balance in Experiments. *Annals of Statistics* 40 (2): 1263-1282.

OECD (2016). *The Survey of Adult Skills: Reader's Companion, Second Edition*. OECD Skills Studies. Paris: OECD Publishing.

Oreopoulos, Philip, Robert S. Brown, Adam M. Lavecchia (2017). Pathways to Education: An Integrated Approach to Helping at-Risk High School Students. *Journal of Political Economy* 125 (4): 947-984.

Piopiunik, Marc, Franziska Kugler, Ludger Woessmann (2017). Einkommenserträge Von Bildungsabschlüssen Im Lebensverlauf: Aktuelle Berechnungen Für Deutschland. *ifo Schnelldienst* 70 (7): 19-30.

Piopiunik, Marc, Guido Schwerdt, Lisa Simon, Ludger Woessmann (2020). Skills, Signals, and Employability: An Experimental Investigation. *European Economic Review* 123: 103374.

Rammstedt, Beatrice (2007). The 10-Item Big Five Inventory. *European Journal of Psychological Assessment* 23 (3): 193-201.

Rammstedt, Beatrice, Oliver P. John (2007). Measuring Personality in One Minute or Less: A 10-Item Short Version of the Big Five Inventory in English and German. *Journal of Research in Personality* 41 (1): 203-212.

Westfall, Peter H., S. Stanley Young (1993). *Resampling-Based Multiple Testing: Examples and Methods for P-Value Adjustment*. New York, NY: Wiley.

Zabal, Anouk, Silke Martin, Beatrice Rammstedt (2016). PIAAC-L Data Collection 2014: Technical Report. Follow-up to PIAAC Germany 2012. GESIS Papers 2016/17.