

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, IWAAEE, 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), CESifo, and ROA, 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), $HI_SES_{ip(t_0)}$:

$$Y_{ipt} = \alpha_0 + \alpha_1 T_{ip} + \alpha_2 T_{ip} \times HI_SES_{ip(t_0)} + \alpha_3 HI_SES_{ip(t_0)} + \mathbf{X}'_{ip(t_0)} \boldsymbol{\alpha}_4 + \mu_p + \epsilon_{ipt} \quad (1)$$

The vector \mathbf{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 μ_p for each matched pair. ϵ_{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 α_1 . The coefficient α_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 $D_{ip} \times \widehat{HI_SES}_{ip(t_0)}$ of the corresponding first stage):

$$D_{ip} = \gamma_0 + \gamma_1 T_{ip} + \gamma_2 T_{ip} \times HI_SES_{ip(t_0)} + \gamma_3 HI_SES_{ip(t_0)} + \mathbf{X}'_{ip(t_0)} \boldsymbol{\gamma}_4 + \zeta_p + \omega_{ipt} \quad (2)$$

$$Y_{ipt} = \phi_0 + \phi_1 \widehat{D}_{ip} + \phi_2 \widehat{D}_{ip} \times \widehat{HI_SES}_{ip(t_0)} + \phi_3 HI_SES_{ip(t_0)} + \mathbf{X}'_{ip(t_0)} \boldsymbol{\phi}_4 + \theta_p + v_{ipt} \quad (3)$$

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 $\mathbf{X}_{i(t_0)}$ in the respective subsample:⁴⁶

$$Y_{it} = \beta_0 + \beta_1 T_i + \mathbf{X}'_{i(t_0)} \boldsymbol{\beta}_2 + \varphi_{it} \quad (4)$$

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 $\mathbf{X}_{i(t_0)}$. This allows us to rewrite equation 4 as:

$$Y_{it} = \beta_0 + \beta_1^{residual} T_i + \sum_k \theta^k M_{it}^k + \mathbf{X}_{i(t_0)}' \boldsymbol{\beta}_2 + \mu_{it} \quad (5)$$

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:

$$M_{it}^k = \delta_0^k + \delta_1^k T_i + \mathbf{X}_{i(t_0)}' \boldsymbol{\delta}_2^k + v_{it} \quad (6)$$

The share of the overall treatment effect that can be attributed to the k^{th} mediator can then be calculated by multiplying the treatment effect on the mediator δ_1^k with the impact of the mediator on the outcome θ^k and dividing by the reduced-form treatment effect on the outcome β_1 :

$$\text{share } M_k = \theta^k \delta_1^k / \beta_1 \quad (7)$$

By the virtue of randomization β_1 and δ_1^k are identified, whereas the analysis has to rely on the arguably strong assumption of conditional independence to identify θ^k .⁴⁷ The assumption implies that any potential unobserved mediator or control subsumed in the error term μ_{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 θ^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 θ^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 Chancel, 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 Reevaluation 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, Anastassiya 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.