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

DISCUSSION PAPER SERIES

IZA DP No. 14097

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

Sven Resnjanskij
Jens Ruhose
Simon Wiederhold
Ludger Woessmann

FEBRUARY 2021

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14097

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

Sven Resnjanskij
ifo Institute at the University of Munich and CESifo

Simon Wiederhold
Catholic University Eichstaett-Ingolstadt, BESH, CESifo and ROA

Jens Ruhose
Kiel University, CESifo and IZA

Ludger Woessmann
University of Munich, ifo Institute, CESifo and IZA

FEBRUARY 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

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

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

www.iza.org

ABSTRACT

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

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 adolescents' later labor-market success: math grades, patience/social skills, and labor-market orientation. For low-SES adolescents, the one-to-one mentoring increases a combined index of the outcomes by 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. 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.

JEL Classification: I24, J24, H52

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

Corresponding author:
Ludger Woessmann
Ifo Institute for Economic Research at the University of Munich
Poschingerstr. 5
81679 Munich
Germany

E-mail: woessmann@ifo.de

* We gratefully acknowledge comments and suggestions from Flavio Cunha, Fabian Kosse, Philip Oreopoulos, and seminar participants at the virtual CESifo Area Conference in the Economics of Education, the Universities of Munich and Bamberg, and the ifo Center for the Economics of Education. We are grateful to Franziska Hampf for exceptional collaboration in the survey execution and data collection and to Hanna Brosch, 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/ Stifterverband für die Deutsche Wissenschaft), 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).

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 with the aim 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 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 oversubscription. After surveying all adolescents before program start, we implemented a pair-

[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)).

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), combining 94.5 percent participation in the follow-up survey and collection of administrative grade information from schools for 95.5 percent of participants.

We investigate 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 for participation. Our baseline specification uses a simple sample split (roughly half and half) based on the number of books at home, but results are confirmed with a broader SES index.

We find that the highly disadvantaged youths benefit strongly from participation in the mentoring program. After one year, 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 0.56 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 higher-SES adolescents is highly significant. Average program effects are significantly positive, but relatively modest in size.

[2] In section 4.2 and Appendix F, 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.

Also for each of the three sub-indices – the cognitive, behavioral, and volitional components – the mentoring program has a significant positive treatment effect for low-SES adolescents, but an insignificant negative effect for higher-SES adolescents. For low-SES adolescents, school grades in math increase by 0.29 standard deviations, with achievement raised throughout the grade distribution. The program increases their index of patience and social skills, as well as its patience sub-component, by 0.44 standard deviations. Effects on the social-skills sub-component – which comprises prosociality, trust, and self-efficacy – are somewhat smaller at 0.22 standard deviations and do not reach statistical significance. The index of labor-market orientation rises by 0.29 standard deviations for low-SES adolescents. This effect is driven by the sub-component of wishing to pursue an apprenticeship,[4] with no significant effect on knowing the desired occupational career. 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.

More detailed analyses confirm the overall pattern. There are significant treatment effects for the subgroup of pupils with a migrant background, who constitute 58 percent of our sample. Treatment effects are significantly positive both for low-SES males and low-SES females, with no significant gender difference. Analysis of the number of program participants at the school and classroom level provide no indication of spillover effects on non-participating peers. Results also prove robust to alternative definitions of the main variables and in alternative samples.

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 one third of the overall treatment effect for low-SES adolescents. In the higher-SES sample, about half of the (small and insignificant) negative treatment effect can be attributed to crowding-out of both participation in social school

[4] For most program participants, a successful transition into professional life would imply completing an apprenticeship, which – compared to no professional qualification – is associated with substantially lower unemployment (4.2 vs. 19.1 percent, Institut für Arbeits- und Berufsforschung (2017)) and 31 percent higher lifetime earnings (Piopiunik, Kugler, and Woessmann (2017)) on the German labor market.

activities and parental involvement, which is not present in the low-SES sample. Descriptive analysis of information on the mentoring relationships in the treatment group indicates that low-SES adolescents are more likely than higher-SES adolescents to view their mentor as helpful for improving performance in school and for solving non-school-related problems. By contrast, there are no relevant SES differences in the frequency, duration, or content of the mentoring meetings. Together, the results from the mechanism analyses suggest that the mentoring is successful only if adolescents lack adult support and that qualitative factors of the mentor-mentee relationships matter more for the effectiveness of mentoring than mere program intensity.

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 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, respectively. Section 4 describes the main variables and section 5 discusses the empirical strategy. Sections 6 and 7 report our main results and additional analyses. Section 8 presents an analysis of mechanisms. Section 9 concludes with considerations of the cost-benefit balance and scalability of the program.

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 2009. 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, 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 about two years (with the second year

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

being optional), 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 students 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. In addition, teachers and principals are free to recommend adolescents who they feel would benefit most from the program. 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 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.

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

[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 is meant to lay the foundation for an effective relationship. 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 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.

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 survey one year after program start (section 3.3).

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); see Appendix B.1 for details). All contacted sites and schools agreed to participate.

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 for an overview and Appendix B.2 for details).[10] Appendix Table A1 lists the participating sites and provides the survey dates as well as numbers of schools and participating adolescents for each site. Randomization was performed directly after the baseline survey in each site, and the program started shortly afterwards. About one year after program start (for each site and cohort), we fielded a follow-up survey to collect outcome data. Consequently, the survey field period for the second cohort ended in June 2019.

[9] These incentives were mainly one of the following: cinema ticket, invitation to a Christmas party, one-day job training, or firm visit. In practice, demand for these incentives was typically very low.

[10] The first cohort also includes two pilot studies fielded in November 2015 and June 2016 (see Appendix B).

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 for a set of eleven baseline variables (see Appendix C for details). In three quarters of the matched pairs, the two pupils attend the same classroom. 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 due to a participant leaving the sample by also

[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.

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 sufficient oversubscription of applicants (twice as many applying pupils as there were 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 Follow-Up Survey

To evaluate the effects of the mentoring program on labor-market prospects, we surveyed adolescents in treatment and control groups again about one year after the baseline survey (see Figure 1 and Appendix Table A2). The follow-up survey was conducted similarly to the baseline, i.e., respondents filled the surveys in 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

[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] Appendix Table A1 provides information on the total and randomized samples in each site. Appendix Table A3 shows that adolescents who could not be included in the randomization are similar to those in the randomized sample. The only differences that are statistically significant at the 5 percent level are in patience and the shares of missing survey observations on math grades. 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).

[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.

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-up. These administrative data are available from pupils' report cards that are stored in the respective schools' archives.

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.

Detailed attrition analysis in Appendix D indicates that (the low level of) attrition is very similar in treatment and control groups and is not selective with respect to observables in the baseline period in the total sample or in the subsamples of low-SES and higher-SES adolescents.

4. Data and Variable Definitions

This section describes how we measure adolescents' socioeconomic background as a potential source of treatment-effect heterogeneity (section 4.1) and how we construct our outcome measures of adolescents' labor-market prospects (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 lower-track schools in disadvantaged neighborhoods in relatively large cities, each of which leads to a disproportionately high share of disadvantaged youths. However,

[18] Results are robust to adding survey mode fixed effects and to restricting the sample to participants who conducted the survey at school (not shown). 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 program does not implement any screening or selection of applying adolescents within the participating schools, leading to rather diverse family backgrounds among participants.

Our preferred measure of adolescents' socioeconomic status (SES) is the number of books at home, a powerful proxy for the social, economic, and educational background of children's families frequently been used in the literature (see Schuetz, Ursprung, and Woessmann (2008)). Compared to other SES indicators such as parental education, occupation, or income, books at home are less prone to measurement error and missing responses due to children misreporting or not knowing their parents' information. For example, 40 (32) percent of participants in our survey do not report the education level of their father (mother), whereas all provide information on the number of books in their home (see Appendix E.3 for details). Moreover, the same educational degree may reflect different education outcomes as parents finished their education at different points in time and often in different countries. Still, in Appendix E.3 we show that our results are robust to a more encompassing measure of SES background comprised of books at home, parental education, and employment status.

Appendix Table A4 compares summary statistics of our sample to the general population of adolescents observed in the representative PISA survey.[19] Columns 1 and 2 show that respondents in our study live in households with far fewer books than the average adolescent in Germany. The share of respondents with at most 25 books at home is 47 percent in our sample, compared to only 23 percent in PISA, whereas respondents with more than 100 books are underrepresented in our sample (28 percent vs. 49 percent in PISA). Moreover, 58 percent of our respondents have a migrant background (i.e., respondent or at least one parent not born in Germany), which is more than twice the migrant share in the average youth population (28 percent). At the same time, 22 (23) percent of respondents have a father (mother) with a university degree, which is as large as (even larger than) the respective share in the average population. Parental employment does not differ systematically from the average population.

To distinguish between highly disadvantaged ("low-SES") and more advantaged ("higher-SES") adolescents in our sample, we use the information on books at home. We define low-SES respondents as those with at most 25 books at home (47 percent of the sample) and higher-SES

[19] The national PISA sample used here is representative of ninth-graders. We use the 2012 rather than the 2015 PISA wave because it includes more variables that allow for a characterization of respondents' SES (e.g., parental employment). Note that 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.

respondents as those with more than 25 books at home (53 percent). Columns 3-8 of Appendix Table A4 show that low-SES respondents can indeed be considered as highly disadvantaged along several dimensions. Compared to their higher-SES counterparts, parents of low-SES adolescents are more likely to have a migrant background (72 vs. 45 percent) and less likely to have a university education (e.g., 12 vs. 33 percent for mothers) and to be employed (e.g., 71 vs. 81 percent for fathers). Parents of low-SES respondents also support their children less in school matters (not shown). Notably, low-SES respondents also tend to be disadvantaged compared to similarly defined low-SES respondents in PISA (column 5), which may partly reflect the substantially higher share of migrants in our sample.

When planning the design of the RCT, our hypothesis was that mentoring is mainly successful for highly disadvantaged adolescents who are severely lacking family resources. While the program might also be useful for the 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 mentor crowds out more useful inputs offered by more advantaged families. Therefore, we investigate heterogeneous treatment effects for low-SES and higher-SES adolescents throughout.

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 at the time of the follow-up survey precludes investigation of realized labor-market outcomes.[20] 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 measures of the three components in an overall index of labor-market prospects, but also report results for the three sub-indices and their components. Apart from allowing for an overall assessment of program effectiveness, the aggregation into an outcome index addresses concerns of multiple hypothesis testing by combining all outcome indicators into

[20] As 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 meaningful evaluation of labor-market success will not be available until many years after the follow-up survey.

one measure and improves the statistical power to detect effects (Kling, Liebman, and Katz (2007); Anderson (2008); Heller et al. (2017)). Following the procedure of Kling, Liebman, and Katz (2007), the overall index, the sub-indices, and the separate outcome variables that combine multiple items from the survey questionnaire are all constructed as an equally weighted average of the *z*-scores of the included items. Scores are computed by subtracting the control-group mean and dividing by the control-group standard deviation, separately by survey round.[21]

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 PIAAC and PIAAC-L for Germany, we show in Appendix F.2 that math grades in school are strong predictors of later cognitive skills in numeracy, literacy, and mastering information and communication technology (ICT) in adulthood (see Appendix Table F1). Prior research suggests that these cognitive skills – especially numeracy – are important determinants of individuals' wages and employment on the labor market, and particularly so in Germany (e.g., 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 (Appendix Table F1). Moreover, since more than half of our sample consists of respondents with migrant background, language difficulties may introduce measurement error by influencing school performance in language classes.[22] 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.[23] Grades are directly comparable between treatment and control pupils in each matched pair because the two pupils in each pair attend the

[21] 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)).

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

[23] 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.

same school and, in three quarters of cases, even the same classroom.[24] Grades are standardized and the usual German ordering is reversed so that higher values indicate better outcomes.[25]

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-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.[26]

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. (2020)) and labor-market success in adulthood (Golsteyn, Grönqvist, and Lindahl (2014)).[27] Other concepts such as grit, conscientiousness, perseverance, and commitment, which are likely related to patience, have also been shown to be relevant for labor-market success (see Almlund et al. (2011) for an overview). In our analysis of the German PIAAC/PIAAC-L data, higher levels of grit are associated with lower employment risk and higher wages particularly for low-SES individuals (Appendix F.3). In addition, higher levels of patience may increase the likelihood that adolescents continue and successfully complete an apprenticeship, particularly so for low-SES individuals who are much more likely to quit an apprenticeship than their higher-SES counterparts (Appendix F.4).

Recent evidence also highlights the 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 held about others' trustworthiness – for which evidence for its relevance for individuals' labor-market outcomes is scarcer, with Butler, Giuliano, and Guiso (2016) as a noticeable exception. In the

[24] 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 7.3 below).

[25] 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.

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

[27] At the macroeconomic level, countries with more patient populations are wealthier and grow faster (Galor and Özak (2016); Dohmen et al. (2019)).

German PIAAC/PIAAC-L data, we find that trust is positively associated with employment prospects and wages (Appendix F.3).[28]

Measurement. We use survey responses to measure patience and social skills, relying on established taxonomies and survey items (see Appendix Table A5 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)).

c) Volitional Component: Labor-Market Orientation

Relevance. The third component of our index of labor-market prospects is the volitional component of labor-market orientation. 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)). There is a large gap in the failure to obtain at least an apprenticeship-level professional qualification between low-SES individuals (20 percent) and higher-SES individuals (5 percent), which emerges already early in the career and is highly persistent (Appendix F.4). 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 lower-track secondary schools (Appendix F.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 variable measuring the wish to get an apprenticeship takes a value of one if respondents state that they would like to do

[28] Aggregate evidence supports a positive relationship between trust and income at the country level (Knack and Keefer (1997); Algan and Cahuc (2010)).

an apprenticeship after finishing school and zero otherwise.[29] Knowledge about the future career 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 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).[30]

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.[31] Thus, the randomization procedure achieves balancing in the full sample and in both SES subsamples.[32]

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

[30] 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.

[31] None of the 29 *F*-tests is significant at the 5 percent level, and only two (those referring to administrative math grades ($p = 0.092$) and survey-based English grades ($p = 0.093$)) are barely significant at 10 percent.

[32] Appendix Table A6 provides comparisons of the baseline variables between treatment and control groups within the subsamples of low-SES and higher-SES respondents, respectively. Across all three samples (full, low-

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 year after program start). 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 $t-1$), $HI\_SES_{ip(t-1)}$, which takes a value of one if the respondent had more than 25 books at home ("higher-SES") and zero otherwise ("low-SES"):

$$Y_{ipt} = \alpha_0 + \alpha_1 T_{ip} + \alpha_2 T_{ip} \times HI\_SES_{ip(t-1)} + \alpha_3 HI\_SES_{ip(t-1)} + \boldsymbol{X}'_{ip(t-1)}\boldsymbol{\alpha_4} + \mu_p + \epsilon_{ipt} \quad (1)$$

The vector $\boldsymbol{X}$ includes control variables from the baseline survey to improve precision of the estimation. Importantly, regressions control for the pre-treatment observation of the respective outcome variable. Additional pre-treatment control variables are gender, age, and migrant status as demographic variables; paid private teaching and parental homework support as non-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 A5 for details). 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.[33] Take-up is somewhat lower for low-SES

SES, higher-SES), none of the variables differs between treatment and control groups at a significance level of 5 percent, and only one variable (self-efficacy in the higher-SES sample) differs at the 10 percent level.

[33] 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.

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):

$$D_{ip} = \gamma_0 + \gamma_1 T_{ip} + \gamma_2 T_{ip} \times HI\_SES_{ip(t-1)} + \gamma_3 HI\_SES_{ip(t-1)} + \boldsymbol{X}'_{ip(t-1)}\boldsymbol{\gamma_4} + \zeta_p + \omega_{ipt} \quad (2)$$

$$Y_{ipt} = \phi_0 + \phi_1 \widehat{D_{ip}} + \phi_2 D_{ip} \times \widehat{HI\_SES}_{ip(t-1)} + \phi_3 HI\_SES_{ip(t-1)} + \boldsymbol{X}'_{ip(t-1)}\boldsymbol{\phi_4} + \theta_p + \nu_{ipt} \quad (3)$$

Since we randomize at the individual level, we cannot rule out a priori that pupils in the control group benefit from treated peers, which would lead to an underestimation of the program impact. We investigate this possibility of spillover effects in section 7.2 below.

6. Main Results

We begin by estimating the effect of the mentoring program on the summary index of labor-market prospects (section 6.1). We then estimate treatment effects on each of the three sub-indices – cognitive, behavioral, and volitional aspects of labor-market prospects – as well as their respective individual components (section 6.2).

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-hand 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 15.3 percent of a standard deviation higher than that of adolescents in the control group, significant at the 10 percent level ($p = 0.089$).

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 55.6 percent of a standard deviation higher than for low-SES adolescents in the control group ($p < 0.001$; see middle panel of Figure 2). In fact, after participating in the program for one year, the index of labor-market prospects of low-SES adolescents is similar to that of higher-SES adolescents in the control group (right-hand panel).

Thus, program participation fully closes the SES gap in labor-market prospects in our sample (indicated by the difference in control-group means between the higher-SES and low-SES subsamples) for low-SES adolescents. By contrast, the relatively more advantaged adolescents do not benefit from the program. If anything, they tend to be negatively affected, but the treatment effect is relatively small at 19.2 percent of a standard deviation and not statistically significant at conventional levels.

Table 2 shows the corresponding regression estimates of the ITT effects for the two subsamples of low-SES and higher-SES adolescents.[34] Treatment effects remain very similar across the specifications. Results of the unconditional model in column 1 are remarkably close to those shown in Figure 2, which are based on a specification that make use of the baseline survey information and matched-pair design.[35] Column 2 adds controls for pre-treatment values of the outcome, column 3 a full set of fixed effects for the randomization pairs obtained from pair-wise matching, and column 4 additional controls for individual characteristics.[36] In all specifications, there is a large treatment effect for low-SES adolescents, which is highly significant both when inference is based on robust standard errors and when using randomization inference. 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, albeit relatively small and statistically insignificant across specifications.[37]

Estimates of the TOT effect of the mentoring program for those adolescents who actually took up the program are shown in column 5. With non-compliance, the TOT effect is larger than

[34] Appendix Table A7 provides analogous estimates for the average program effect, not distinguishing between low- and higher-SES participants (column 4 corresponds to the left-hand panel in Figure 2). The average treatment effect estimates are positive across specifications and reach statistical significance at conventional levels when the pre-treatment outcome is controlled for. Average treatment effects for the three sub-indices are also positive but fail to capture statistical significance (not shown).

[35] Appendix Figure A1 shows the unconditional treatment effects on the entire distributions of labor-market prospects. Analogously to the comparison of mean effects in Figure 2, 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 ($p = 0.828$).

[36] Appendix Table A8 reports coefficient estimates for the covariate characteristics. Due to the inclusion of randomization-pair fixed effects and since most control variables were used as balancing variables in the pair-wise matching, almost all controls are insignificant.

[37] A sample split by SES background shows very similar (and even slightly stronger) treatment effects compared to the interaction specification (Appendix Table A9). Results are also similar (and again slightly stronger) when dropping pairs in which matched partners have a different SES (columns 3 and 4 of Appendix Table A9).

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 68.4 percent of a standard deviation.[38]

The positive program impact on individuals' labor-market prospects is mirrored in a positive effect on overall life satisfaction. Low-SES youths in the treatment group are 22.4 percent more likely to be satisfied with their lives than their counterparts in the control group (column 1 of Appendix Table A11). Such an increase in overall life satisfaction may materialize through an increase in participants' labor-market prospects or through other factors positively affected by the program. There is no significant treatment effect on life satisfaction for higher-SES youths.

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

In the following, we separate the index of labor-market prospects into its three sub-indices, as well as their respective components. 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.[39]

Math Achievement in School. Table 3 reports program effects on administrative math grades in school. In column 1, 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 29.4 percent of a standard deviation, closing more than half of the SES achievement gap. The mentoring program does not significantly affect the math achievement of higher-SES adolescents.[40]

A more fine-grained picture on math achievement is shown in columns 2-5 which use indicators of achieving at least a specific math grade ("very good," "good," "satisfied," and "pass," respectively). Results indicate that the mentoring program raises achievement throughout the grade distribution. For instance, column 2 indicates that for low-SES adolescents, program participation increases the likelihood to achieve a "very good" math grade by 12.9 percentage

[38] Appendix Table A10 shows TOT results for the sub-indices.

[39] 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 multiple hypotheses corrections, the List, Shaikh, and Xu (2019) procedure based on Romano and Wolf (2010) and the Westfall and Young (1993) procedure. Appendix Table A12 reports the adjusted *p*-values, which range from 0.008 for the patience and social skills index using the Westfall-Young correction to 0.087 for the labor-market orientation index using the List-Shaikh-Xu correction.

[40] There is also an insignificant positive effect on low-SES adolescents' satisfaction with their math performance (Appendix Table A11, column 2); the fact that this effect is weaker than the effect on actual performance may indicate that treated adolescents may have had even higher aspirations.

points (two thirds of the control-group mean indicated at the bottom of the table), closing 58 percent of the SES gap in this outcome.

Patience and Social Skills. Table 4 shows program effects on patience and social skills. Column 1 uses a summary measure that combines patience as a main dimension of non-cognitive skills and an index of social skills. The summary measure is an equally-weighted average of *z*-scores of its two components. For low-SES adolescents, program participation increases the index of patience and social skills by 43.9 percent of a standard deviation, fully closing the SES gap in this outcome for low-SES participants. 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 2). The program effect of 44.1 percent of a standard deviation for low-SES adolescents exceeds the control-group gap in patience between higher-SES and low-SES adolescents (27 percent of a standard deviation).

The treatment effect on the social-skills index is also positive for low-SES adolescents, but smaller (21.7 percent of a standard deviation) and not statistically significant at conventional levels (column 3). 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 (columns 4-6).[41]

Labor-Market Orientation. The mentoring program also raises the labor-market orientation of low-SES adolescents. Column 1 of Table 5 shows that treatment increases the index of labor-market orientation of low-SES adolescents by 29.1 percent of a standard deviation. Program effects on the labor-market orientation index of higher-SES adolescents are close to zero. Looking at the separate components of the index, there is a sizeable treatment effect on highly disadvantaged youths' wish to get an apprenticeship after school (column 2).[42] By contrast, the program does not affect adolescents' knowledge about their future career (column 3).

[41] In line with the relatively weak effects on social skills, Appendix Table A13 shows that the mentoring program has no effect on social capital as measured by volunteering, the number of friends, and the frequency of meeting friends. For low-SES adolescents, program participation neither affects a series of measures of school-related social capital (Appendix Table A14). 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 8.1 below).

[42] 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 (Appendix Table A15, column 2).

We interpret this evidence as showing 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). This interpretation is also consistent with results on various dimensions of satisfaction (Appendix Table A11): Low-SES youths in the treatment group are more satisfied with their lives (column 1) and their current belongings (column 5) than their counterparts in the control group, suggesting that the program makes highly disadvantaged adolescents focus on what they can realistically achieve and appreciate what they already possess.

7. Additional Analyses

This section provides additional evidence on treatment-effect heterogeneity (section 7.1), shows that spillover effects are unlikely to be a major threat to identification (section 7.2), and demonstrates that results are robust to alternative variable definitions and samples (section 7.3).

7.1 Additional Analysis of Effect Heterogeneity

Our baseline specification shows important effect heterogeneity between low-SES and higher-SES adolescents. Here, we investigate treatment effects for subgroups that are also often associated with a low socioeconomic background: migrants and single-parent families. We also investigate effect heterogeneity by gender and whether the program impact is more detrimental for individuals in the highest part of the SES distribution.

The first additional dimension of effect heterogeneity indicates that treatment effects are larger for migrants, in particular first-generation migrants, than for natives. In our sample, as well as in Germany overall (e.g., Algan et al. (2010)), migrants tend to be overrepresented in the low-SES group. In line with this, we find significantly positive treatment effects for migrant pupils (although smaller than for low-SES pupils) and no effect for non-migrants on average (see Appendix E.1 for details). The mentoring program fully closes the native-migrant gap in labor-market prospects. In contrast to the results for low-SES adolescents, the program achieves only very modest impacts on math performance as well as on patience and social skills for migrants. Instead, it strongly increases migrants' labor-market orientation, where migrants show a large gap to natives. The program is particularly successful in improving the labor-market prospects of first-generation migrants (adolescent born abroad, 22 percent of migrants) rather than second-generation migrants (adolescent born in Germany, at least one parent born abroad).

By contrast, there is no strong effect heterogeneity by single-parenthood status. Single parenthood is another characteristic that is often associated with low socioeconomic status (e.g., Kosse et al. (2020)). Overall, 25 percent of the adolescents in our sample live with only one parent, substantially more than in the overall adolescent population (14 percent in PISA). Consistent with the view that single parents can often provide only limited resources, the treatment effect is stronger for adolescents from single-parent families than for those from two-parent families, albeit not significantly so (see Appendix E.2 for details).

We also investigate effect heterogeneity by gender. The mentoring program shows significant positive effects for both low-SES males and low-SES females, and the difference between the genders is not statistically significant (see Appendix Table A16).

To further investigate the negative (albeit insignificant) treatment effect for higher-SES adolescents, we test whether the program impact is more detrimental for individuals in the highest part of the SES distribution. We split the indicator of higher-SES background into "medium SES" (26-200 books at home) and "high SES" (more than 200 books at home). In line with the mission of the program, high-SES individuals are less represented in the program than their medium-SES counterparts (17 vs. 36 percent). Appendix Table A17 shows that program effects are not systematically or significantly more negative for high-SES than for medium-SES adolescents. Moreover, program effects are insignificant in both subgroups. The relatively small number of high-SES adolescents in the sample and the fact that treatment effect sizes are not systematically different between medium-SES and high-SES adolescents provide a justification for pooling medium-SES and high-SES into one group in the main analysis.

7.2 Tests for Spillover Effects on Non-Participating Peers

If the mentoring program was to exert spillover effects on non-treated youths who attend the same school or classroom as treated youth, estimated treatment effects would be attenuated. To get an idea on the possible relevance of spillover effects for our results, we perform two types of non-experimental analyses that relate control-group performance to measures of the extent of program participation in a school or classroom (see Heller et al. (2017)). We calculate three versions of the measure of program participation, each leading to very similar results: the total number of treated pupils in the school cohort, the total number of treated pupils in the classroom, and the share of treated pupils in the classroom. We standardize the program participation variables for comparability and refer to them as *Treated pupils*.

The first analysis interacts the measures of *Treated pupils* with the indicator variables for treatment and higher-SES (Appendix Table A18).[43] The existence of spillover effects from treated pupils to control pupils would imply that *Treated pupils* and/or the interaction of *Treated pupils* and *Higher-SES* are significantly positive, since being surrounded by more treated pupils would positively affect the outcomes of pupils in the control group. However, the respective coefficients are small and statistically insignificant. Treatment effects also do not differ by the number of treated peers in the school or classroom.

The second analysis restricts the analysis to the control group and investigates whether control-group pupils perform better if they are in schools or classrooms with more treated pupils. As shown in Appendix Table A19, this is not the case, neither for the full sample (odd columns) nor when distinguishing between low-SES and higher-SES pupils (even columns). If at all, only higher-SES control pupils seem to benefit from being surrounded by more treated pupils, but effect sizes are small and statistically insignificant.

As both the number of participating schools and the sample size in each school are rather small, the statistical power of this spillover analysis is quite limited. However, if pupils in the control group profit from the mentoring program without having participated through peer spillovers or other mechanisms (or if negative spillovers from interacting with non-participating peers undermine effects of the program), our treatment effect estimates would understate the effect of offering the mentoring program at a larger scale.

7.3 Further Robustness Analyses

This section shows that the baseline results are robust to a series of alternative definitions of the main variables and alternative samples.

For the conceptional and data-quality reasons described in section 4.1, our main SES measure is based on books at home. However, we can use information on parental education and employment status to construct a more encompassing index of SES (see Appendix E.3 for details). Qualitative results for this broader SES index are very similar to our baseline results (Appendix Table E7). The slightly lower precision of the estimates likely stems from increased measurement error due to missing values on the additional SES dimensions.

[43] Since our randomization procedure matches on class, three quarters of treatment and control youths within a matched pair are in the same class. Thus, there would be only little variation in the participation variables to exploit if we conditioned on randomization-pair fixed effects. To use as much variation as possible, Appendix Table A18 controls neither for randomization-pair fixed effects nor for individual background variables.

While we prefer a lean specification of the sub-index of labor-market orientation, some additional pieces of information from the questionnaire can be used to construct a broader index. In particular, we elicited the following information related to adolescents' labor-market orientation: already applied for apprenticeships or plan to apply during the school year; participation in job-preparation events, career entry support, or job coaching; and importance of job agencies and vocational preparation at school as sources of information for career choice. Column 1 of Appendix Table A20 shows that a broader labor-market orientation index that includes these additional items leads to very similar results compared to the leaner index. In fact, treatment effects for low-SES individuals become even larger and are more precisely estimated than in the main specification. For low-SES adolescents, the point estimates of program participation are positive for each individual component of the labor-market orientation index; however, among the newly included items, none of the treatment effects captures statistical significance (columns 4-6 of Appendix Table A20). Results are also very similar to those in the main specification when we use the broader index of labor-market orientation to construct the summary measure of labor-market prospects (Appendix Table A21).

Results are also robust to separately excluding individual mentoring sites (Appendix Table A22). As the number of baseline observations differs by mentoring site, ranging from just 6 observations in Leipzig to 55 observations in Aachen (see Appendix Table A1), it is reassuring that the program effect is not driven by any specific site. The size of the estimated treatment effect varies somewhat depending on which site is excluded, but this variation appears unrelated to the number of site-specific observations.[44] In further analysis, we also estimate site-specific treatment effects by adding a triple interaction between treatment, SES, and indicators for each specific mentoring site. The estimated treatment effect for low-SES adolescents is positive for *each* individual site, and in seven of the nine sites the estimated point estimate is larger than the average effect found in our baseline model (not shown).

In three quarters of the matched pairs, both pupils in the pair attend the same classroom in their school. Results are very similar to the full-sample results when we restrict the sample to pairs where both pupils attend the same classroom (column 2 of Appendix Table A23). Although less precisely estimated, the treatment effect on labor-market prospects is also significant in the

[44] The site-level correlation between the estimated program effect when a specific site is excluded and the number of remaining observations is -0.28 for low-SES adolescents and -0.25 for higher-SES adolescents.

small subsample of pairs where the two pupils attend different classrooms (column 3). Because of the direct comparability of math grades within a classroom, it is also reassuring that the treatment effect on math grades is robust – and in fact even larger than in the full sample – in the subsample of same-classroom pairs (column 2 of Appendix Table A24), whereas it does not show up in the small subsample of pupil pairs not sharing the same classroom (column 3).

8. Analysis of Mechanisms

This section studies a range of potential channels that might underlie the treatment effect of the mentoring program. We show that in a mediation analysis, a considerable share of the low-SES treatment effect can be attributed to having an attachment figure who provides guidance for the future (section 8.1). We also present suggestive evidence that qualitative differences in the mentoring relationship can account for treatment effect heterogeneity by SES (section 8.2).[45]

8.1 Mediation Analysis

The aim of the mediation analysis is to better understand the underlying mechanisms through which the mentoring program affects the labor-market prospects of adolescents. Our analysis follows the approach developed in Heckman, Pinto, and Savelyev (2013) and Heckman and Pinto (2015) (see also Oreopoulos, Brown, and Lavecchia (2017)). Based on the assumption that the outcome can be expressed as a linear combination of mediator variables and baseline demographic controls, the mediation analysis provides a decomposition of the overall treatment effect into shares attributed to different mediators (see Appendix G.1 for details).

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 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 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:

[45] 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.

schools, future orientation, and occupational orientation. To act as mediators, the respective factors – elicited for both treatment and control groups in the background questionnaires – must be significantly affected by the treatment and must be related to the outcome.

The first mediator captures whether, as part of developing a career vision, the mentoring is successful in making mentees perceive schooling as useful for a later job. Using this mediator as dependent variable in our baseline specification (equation 1), program participation indeed increases the extent to which low-SES adolescents, but not higher-SES adolescents, agree that material learnt in school is useful for future jobs (column 1 of Appendix Table G1).[46]

The second mediator captures whether the treatment successfully establishes the mentor as an attachment figure for talking about the future. In the background questionnaire, adolescents report with whom they talk about their future. Program participation raises the likelihood that low-SES (as well as higher-SES) adolescents mention a mentor or coach as a person with whom they talk about their future (column 1 of Appendix Table G2).[47] By contrast, the treatment does not significantly affect the extent to which low-SES adolescents mention other people – parents, siblings, other relatives, friends, teachers, or others – as attachment figures with whom they talk about their future (columns 2-7).

The third mediator captures whether mentors are important for providing information about occupational choice. The treatment significantly increases the likelihood that low-SES (and higher-SES) adolescents consider a mentor or coach as an important source of information for job choice (column 1 of Appendix Table G3).[48] Again, there is no significant treatment effect on the likelihood of receiving important job information from other people, namely family, friends, school, employment agency, or media (columns 2-6).

Figure 3 shows the results of the mediation analysis that considers these three mediators in explaining the effect of the mentoring program for low-SES adolescents (see Appendix G.2 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 schools as useful

[46] The treatment also slightly increases the extent to which low-SES adolescents perceive good grades and recognition by teachers as important, although not significantly so (columns 2-3 of Appendix Table G1).

[47] On average, 43 percent of treated adolescents and 5 percent of control adolescents mention a mentor or coach as an attachment figure for talking about their future.

[48] On average, 59 percent of treated adolescents and 29 percent of control adolescents mention a mentor or coach as someone whom they refer to in order to receive occupational information.

for later jobs account for 5 percent of the overall treatment effect, talking with the mentor about the future for 29 percent, and considering the mentor as an important source of information for job choice for 17 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 34 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 3 provides equivalent decompositions for each of the three components of the index of labor-market prospects. The combined mediators account for between 31 and 55 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 schools as useful for jobs does not 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.

A similar mediation analysis for the higher-SES adolescents indicates that their (small and insignificant) negative treatment effect can partly be attributed to a crowding-out of in-school social activities and of parental attachment (see Appendix G.3 for details). Expectedly, the set of mediators considered in the low-SES analysis does a poor job in explaining the negative higher-SES treatment effect. Instead, we consider mediators that are significantly negatively affected by the treatment in the higher-SES sample. The time that mentees spend with the mentors may in principle crowd out participation in other useful activities as well as parental support and attachment.[49] Indeed, for higher-SES adolescents we find that the mentoring program leads to a

[49] For low-SES adolescents, there is no indication of crowding-out effects of the treatment on social activities or parental support. There are no treatment effects on general social capital (e.g., volunteering and meeting friends, Appendix Table A13), school-related social capital (e.g., acting as class representative or participating in the school theater group, Appendix Table A14), parental homework support, or paid private teaching (columns 4 and 5 of

reduction in school-related social activities (column 1 of Appendix Table A14), in the perceived importance of good grades (column 2 of Appendix Table G1), and in the likelihood that parents act as attachment figures with whom higher-SES adolescents talk about their future (column 2 of Appendix Table G2).[50] Together, these three factors can account for about half of the small negative higher-SES treatment effect in a mediation analysis, with the crowding-out of social activities in school as the dominant channel (column 13 of Appendix Table G5).

8.2 Evidence on the Mentoring Relationships

We can obtain additional insight into the channels that may be responsible for successful mentoring from a descriptive analysis of the characteristics of the mentoring relationships. In the follow-up questionnaire, we elicited information on the stability, intensity, and content of the mentoring relationships from the adolescents in the treatment group. Apart from offering a glimpse into the relationships between mentors and mentees, these data allow us to compare the mentoring relationships between low-SES and higher-SES mentees to better understand why the mentoring program achieves its intended impact for low-SES adolescents, but not for higher-SES adolescents. Appendix Table A25 reports the various characteristics of the mentoring relationships for the full sample and separately for low-SES and higher-SES respondents.

Several qualitative measures of the nature of the mentoring relationship differ significantly by mentees' SES, contributing to our understanding of the heterogeneous treatment effects (panel A). Strikingly, 28 percent of low-SES mentees think that their school performance increased due to their mentors, twice as many as among higher-SES mentees (14 percent, difference significant at $p < 0.05$). Low-SES mentees are also more likely to consider their mentors as helpful in tackling problems outside school than higher-SES mentees (38 vs. 23 percent, $p < 0.05$). In addition, low-SES mentees are more likely to view their mentors as a role model than higher-SES mentees (32 vs. 22 percent, $p = 0.153$). Intriguingly, overall satisfaction with the mentoring relationship does not differ significantly by SES, and parental approval of the

Appendix Table G1). Low-SES adolescents also do not experience a reduction in viewing parents as attachment figures to talk about their future (column 2 of Appendix Table G2) or in viewing their family as a source of job information (column 2 of Appendix Table G3). At the same time, these results also indicate that the program effect does not materialize through significant improvements in the family situation (e.g., by mentors influencing parents' school-related support or educational aspirations). This is in line with the finding that adolescents' satisfaction with family, friends, or school does not improve through the program (columns 6-8 of Appendix Table A11).

[50] Furthermore, for higher-SES adolescents there are non-significant reductions in the perceived usefulness of school for future jobs, in the perceived importance of the recognition by teachers, and in parental homework support (columns 1, 3, and 4 of Appendix Table G1).

mentoring relationship is actually *lower* for low-SES mentees (46 percent) than for higher-SES mentees (61 percent, $p < 0.10$). These comparisons suggest that several qualitative factors of mentors' help with topics within and outside school distinguish the average mentoring relationship between low-SES and higher-SES adolescents.

By contrast, the analysis suggests that the heterogeneity in treatment effects by SES is unlikely to be due to SES differences in the continuation of the relationship or in the frequency, duration, or content of the meetings. First, low-SES adolescents (56 percent) are in fact *less* likely to still have a mentoring relationship one year after program start than higher-SES adolescents (70 percent, see panel B). This is a combination of a slightly lower propensity to take up the mentoring relationship in the first place (see also section 5.2) and a slightly lower propensity to maintain the relationship once initiated.

Second, there are no pronounced SES differences in the frequency and duration of the meetings (panel C). Half of the mentees in both the low-SES and higher-SES samples report that they meet their mentor in person at least once every month. In addition to face-to-face meetings, mentees report that they are frequently in contact with their mentor in other ways, in particular via social-media channels (such as WhatsApp and Facebook) and text messaging. When also including these communication channels, the share of mentees who meet their mentor at least once per month is *lower* for low-SES adolescents (57 percent) than for higher-SES adolescents (66 percent). In addition, meetings of low-SES mentees are somewhat *shorter* on average (2.9 vs. 3.3 hours). None of these differences are statistically significant, however.

Third, the topics discussed during the meetings are also very similar for low-SES and higher-SES adolescents (panel D). The most relevant topics discussed by mentors and mentees are school issues (66 percent), leisure activities (57 percent), the future in general (57 percent), the occupational and educational future in particular (50 percent), and personal issues (49 percent). The only significant difference is that higher-SES mentees are more likely to talk with their mentors about leisure activities than low-SES mentees (67 vs. 46 percent).

Overall, the descriptive analysis of characteristics of the mentoring relationships suggests that the fact that the mentoring program helped low-SES adolescents but not higher-SES adolescents is related to qualitative factors of the mentoring relationships, but not to the frequency, length, or content of the meetings between mentors and mentees.

9. Conclusion

Our results suggest that mentoring programs can successfully improve the future labor-market opportunities of highly disadvantaged youths. For low-SES adolescents, the mentoring program that we study increases a summary measure of labor-market prospects 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 by the program. 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. Our mediation analysis indicates that aspects of the mentor-mentee relationship that help low-SES adolescents develop a career vision, in particular guidance by the mentors for their future, can account for about one third of the overall program effect.

By contrast, the program does not significantly affect higher-SES adolescents. 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 school-related and non-school-related problems, and program participation may even crowd out their social school activities and parental attachment.

A 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. Given the large program effect, the projected gain in discounted lifetime earnings amounts to 23,500 EUR for low-SES adolescents (see Appendix H 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 15-to-1 for an untargeted program to 31-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 degrees of 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.

This raises the question of scalability of successful mentoring programs. There are two aspects to this. First, the strong heterogeneity of results by SES suggests that to have impact,

scaling should focus on those youths who really lack family support. Other adolescents 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. Second, in several regards the program – as well as the design of the field experiment – are geared to show scalability beyond one specific location. The program is organized as a nationwide franchise with a small central holding and mostly self-governing local sites. The system has shown to be able to grow from one to over forty locations within just ten years. What is more, the RCT was not restricted to one or two selected sites, but administered in 10 locations and 19 schools across Germany, ensuring that treatment effects are not driven by any specific location. As a limiting factor, the program so far 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.

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Dohmen, Thomas, Benjamin Enke, Armin Falk, David Huffman, Uwe Sunde (2019). Patience and Comparative Development. Mimeo.

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.

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.

Galor, Oded, Ömer Özak (2016). The Agricultural Origins of Time Preference. *American Economic Review* 106 (10): 3064-3103.

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 (2020). The Effect of Mentoring on School Attendance and Academic Outcomes: A Randomized Evaluation of the Check & Connect Program. NBER Working Paper 27661. Cambridge, MA: National Bureau of Economic Research.

Hanushek, Eric A., Lavinia Kinne, Philipp Lergetporer, Ludger Woessmann (2020). Culture and Student Achievement: The Intertwined Roles of Patience and Risk-Taking. NBER Working Paper No. 27484. Cambridge, MA: National Bureau of Economic Research.

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Rodríguez-Planas, Núria (2012). Longer-Term Impacts of Mentoring, Educational Services, and Learning Incentives: Evidence from a Randomized Trial in the United States. *American Economic Journal: Applied Economics* 4 (4): 121-139.

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

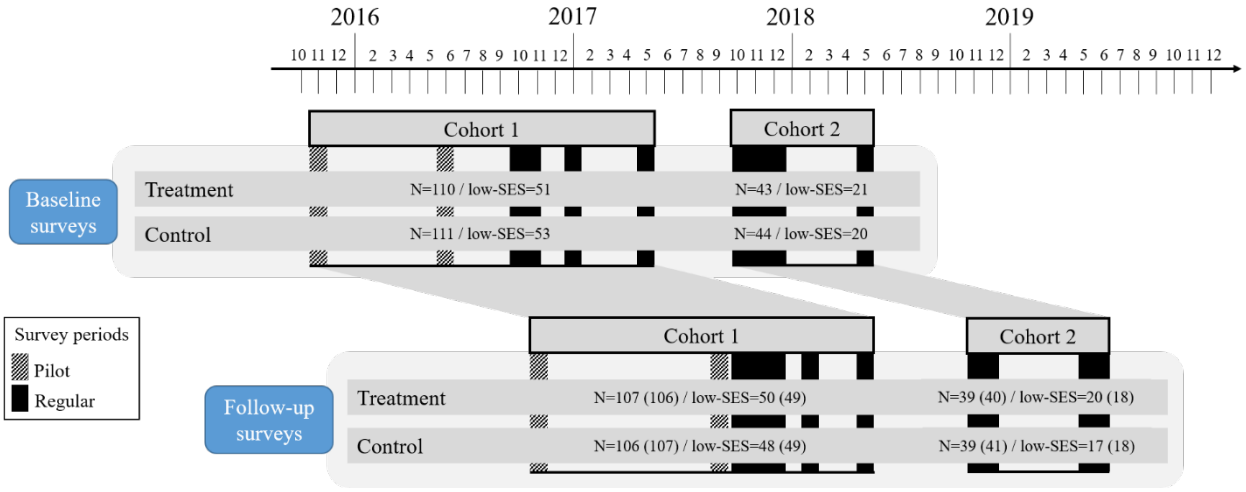
Romano, Joseph P., Michael Wolf (2010). Balanced Control of Generalized Error Rates. *Annals of Statistics* 38 (1): 598-633.

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

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

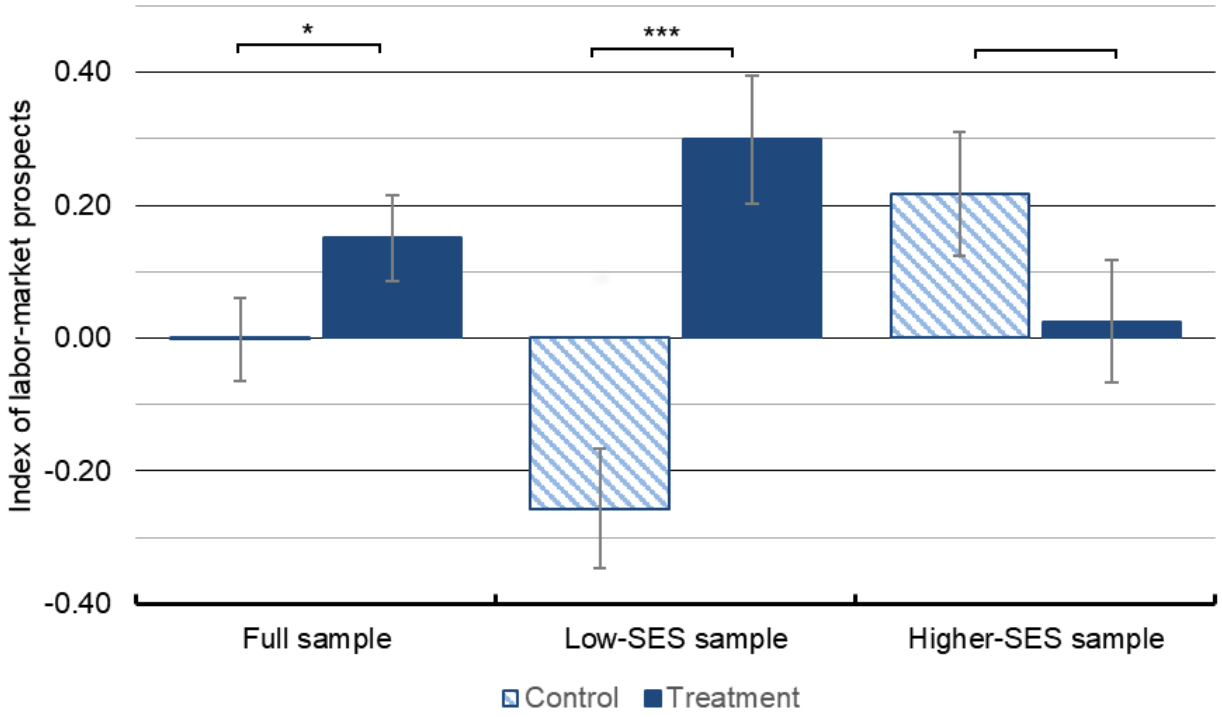
Figures and Tables

Figure 1: Timeline of the Surveys



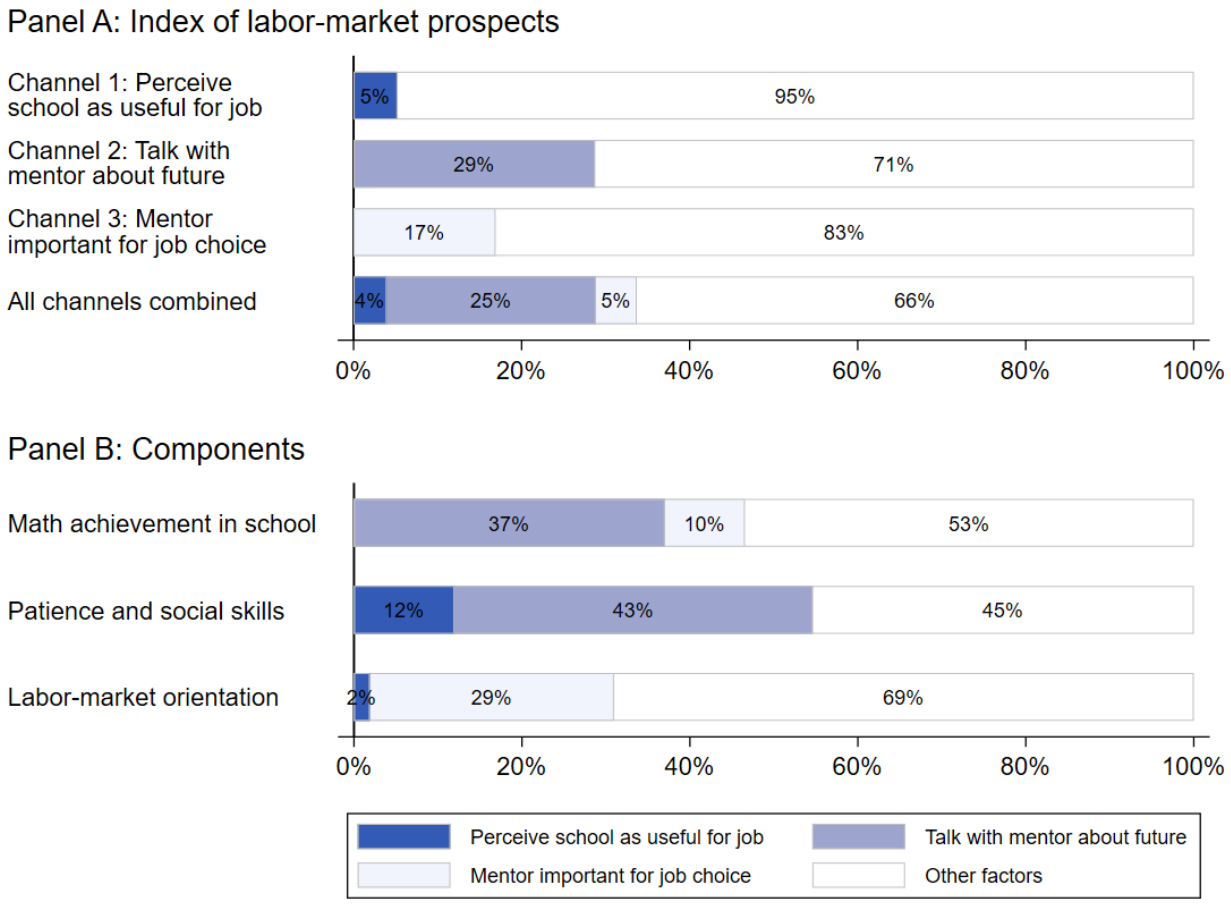
Notes: Figure shows data collection and sample sizes of the randomized sample of the evaluation. Sampling periods, which differ by mentoring site and cohort, are indicated by shaded bars for the pilot surveys and by solid bars for the regular surveys. Treatment started shortly after the baseline survey in each mentoring site. Dates and sample sizes by mentoring site and cohort are shown in Appendix Tables A1 and A2 for the baseline and follow-up surveys, respectively.

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



Notes: Figure shows the intention-to-treat effects (ITT) of the mentoring program on the index of labor-market prospects, separately for all respondents (left panel), low-SES respondents (those with at most 25 books at home at baseline) (middle panel), and higher-SES respondents (those with more than 25 books at home) (right panel). See specification in column 4 of Table 2 for details. The index of labor-market prospects is an equally weighted average of *z*-scores of three components: administrative math grade (reversed), patience and social skills index, and labor-market orientation index. Calculation of each *z*-score subtracts the score's control-group mean and divides by the control-group standard deviation. Error bars show robust standard errors. Significance levels of differences: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure 3: Share of Treatment Effect for Low-SES Adolescents Attributed to Mediators



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

Table 1: Balancing

	Control Mean (1)	Treatment Mean (2)	Difference *p*-value (3)	Difference by SES *p*-value (4)	Observations (5)
A. Outcome variables at baseline					
Overall index	0.00	-0.09	0.433	0.831	308
Components					
Math grade (administrative)	0.00	0.02	0.889	0.092	218
Math grade (admin.) missing dummy	0.28	0.30	0.747	0.885	308
Patience and social skills index	0.00	-0.07	0.548	0.793	308
Labor-market orientation index	0.00	-0.09	0.424	0.415	307
B. Components of outcome variables at baseline					
Patience and social skills index					
Patience	0.00	-0.02	0.891	0.449	308
Social skills index	0.00	-0.09	0.402	0.680	308
Prosociality	0.00	0.01	0.897	0.845	308
Trust	0.00	-0.05	0.665	0.917	307
Self-efficacy	0.00	-0.15	0.158	0.592	308
Labor-market orientation index					
Wants apprenticeship after school	0.36	0.37	0.889	0.836	267
Knows future career	0.00	-0.16	0.156	0.282	307
C. Matching and balancing variables for randomization at baseline					
Male	0.43	0.44	0.921	0.634	308
Age	13.99	13.97	0.851	0.705	308
Migrant	0.59	0.57	0.744	0.710	308
Books at home	1.73	1.67	0.461	0.104	308
Math grade (survey)	1.71	1.73	0.806	0.742	261
Math grade (survey) missing dummy	0.14	0.16	0.602	0.436	308
German grade (survey)	1.73	1.71	0.751	0.431	258
German grade (survey) missing dummy	0.15	0.17	0.721	0.376	308
English grade (survey)	1.79	1.83	0.626	0.093	258
English grade (survey) missing dummy	0.15	0.17	0.721	0.397	308
Received paid private teaching	0.18	0.21	0.529	0.745	308
Parental homework support	2.81	2.71	0.368	0.776	307
Big-5: Conscientiousness	3.35	3.26	0.327	0.132	308
Big-5: Neuroticism	2.91	2.98	0.413	0.729	308
D. Further control variables at baseline					
Big-5: Openness	3.41	3.51	0.337	0.421	308
Big-5: Extraversion	3.31	3.35	0.610	0.704	308
Big-5: Agreeableness	3.50	3.46	0.704	0.859	307
Higher-SES (>25 books at home)	0.53	0.53	0.995	–	308

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

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

	ITT				TOT
	(1)	(2)	(3)	(4)	(5)
Treatment	0.549***	0.573***	0.545***	0.556***	0.684***
	(0.180)	*(0.144)*	*(0.136)*	*(0.143)*	*(0.170)*
	[0.000]	*[0.000]*	*[0.000]*	*[0.000]*	–
Treatment x Higher-SES	-0.751***	-0.701***	-0.659***	-0.748***	-0.895***
	(0.237)	*(0.197)*	*(0.209)*	*(0.220)*	*(0.251)*
	[0.001]	*[0.001]*	*[0.004]*	*[0.002]*	–
Higher-SES	0.476***	0.431***	0.113	0.182	0.201
	(0.158)	*(0.140)*	*(0.184)*	*(0.195)*	*(0.196)*
Outcome in t_0		0.580***	0.501***	0.459***	0.439***
		(0.053)	*(0.065)*	*(0.078)*	*(0.078)*
Randomization-pair fixed effects	No	No	Yes	Yes	Yes
Covariates	No	No	No	Yes	Yes
Observations	304	304	304	304	304
R^2	0.040	0.325	0.723	0.753	0.754
Kleibergen-Paap *F* statistic					141.50
Treatment effect for Higher-SES	-0.202	-0.128	-0.114	-0.192	-0.211
	(0.154)	*(0.135)*	*(0.137)*	*(0.137)*	*(0.151)*
SES gap			0.476		

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

Table 3: Effect of the Mentoring Program on Math Achievement in School

	Math grade	Dummies of specific math grades			
		"very good"	"good" or better	"satisfied" or better	"pass" or better
	(1)	(2)	(3)	(4)	(5)
Treatment	0.294**	0.129*	0.150*	0.153**	0.042
	(0.142)	*(0.076)*	*(0.083)*	*(0.073)*	*(0.057)*
	[0.034]	*[0.103]*	*[0.088]*	*[0.035]*	*[0.498]*
Treatment x Higher-SES	-0.467**	-0.222*	-0.263*	-0.136	-0.051
	(0.230)	*(0.119)*	*(0.139)*	*(0.121)*	*(0.076)*
	[0.036]	*[0.061]*	*[0.057]*	*[0.217]*	*[0.516]*
Higher-SES	0.283	0.168*	0.156	0.094	-0.014
	(0.196)	*(0.095)*	*(0.101)*	*(0.104)*	*(0.070)*
Outcome in t_0	0.488***	0.580***	0.371***	0.261**	0.135
	(0.100)	*(0.125)*	*(0.114)*	*(0.127)*	*(0.168)*
Randomization-pair fixed effects	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes
Observations	294	294	294	294	294
R^2	0.775	0.653	0.694	0.697	0.680
Treatment effect for Higher-SES	-0.172	-0.093	-0.113	0.017	-0.010
	(0.145)	*(0.071)*	*(0.089)*	*(0.080)*	*(0.042)*
SES gap	0.553	0.222	0.217	0.211	0.080
Control-group mean	0.000	0.195	0.432	0.608	0.865

Notes: Table shows ITT effects of the mentoring program on administrative math grades. Column 1: grades are standardized by subtracting the control-group mean and dividing by the control-group standard deviation; order of grades is reversed so that higher values indicate better outcomes. Columns 2-5: dummies indicating achievement of at least the specified grade. Ordinary least squares estimates. *SES gap* is calculated as the coefficient on higher-SES background in a regression of the respective outcome on the higher-SES indicator in the control-group sample in the follow-up survey. *Control-group mean* indicates the mean of the respective outcome in the control-group sample in the follow-up survey. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in t_0 are included. Robust standard errors in parentheses. Randomization inference (RI) *p*-values in square brackets, obtained from RI with 1,000 permutations, assigning the treatment status randomly within randomization pairs. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: Effect of the Mentoring Program on Patience and Social Skills

	Patience and social skills	Patience	Social skills			
	Index	Index	Index	Components		
				Prosociality	Trust	Self-efficacy
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.439***	0.441**	0.217	0.179	0.095	0.151
	(0.152)	*(0.175)*	*(0.177)*	*(0.181)*	*(0.211)*	*(0.181)*
	[0.003]	*[0.019]*	*[0.240]*	*[0.310]*	*[0.628]*	*[0.408]*
Treatment x Higher-SES	-0.587**	-0.536*	-0.321	-0.296	-0.082	-0.285
	(0.255)	*(0.279)*	*(0.240)*	*(0.266)*	*(0.288)*	*(0.250)*
	[0.018]	*[0.061]*	*[0.207]*	*[0.273]*	*[0.780]*	*[0.266]*
Higher-SES	0.164	0.181	0.070	0.151	-0.187	0.119
	(0.227)	*(0.226)*	*(0.224)*	*(0.212)*	*(0.270)*	*(0.224)*
Outcome in t_0	0.261**	0.262***	0.459***	0.419***	0.373***	0.447***
	(0.103)	*(0.086)*	*(0.117)*	*(0.126)*	*(0.112)*	*(0.094)*
Randomization-pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Observations	291	291	291	290	290	291
R^2	0.695	0.648	0.701	0.730	0.601	0.679
Treatment effect for Higher-SES	-0.148	-0.095	-0.104	-0.118	0.013	-0.134
	(0.174)	*(0.181)*	*(0.140)*	*(0.161)*	*(0.166)*	*(0.142)*
SES gap	0.389	0.270	0.320	0.254	0.077	0.313

Notes: Table shows ITT effects of the mentoring program on patience and social skills. Variables and indices are standardized by subtracting the control-group mean and dividing by the control-group standard deviation. Ordinary least squares estimates. *SES gap* is calculated as the coefficient on higher-SES background in a regression of the respective outcome on the higher-SES indicator in the control-group sample in the follow-up survey. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in t_0 are included. Robust standard errors in parentheses. Randomization inference (RI) *p*-values in square brackets, obtained from RI with 1,000 permutations, assigning the treatment status randomly within randomization pairs. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: Effect of the Mentoring Program on Labor-Market Orientation

	Index	Wants apprenticeship after school	Knows future career
	(1)	(2)	(3)
Treatment	0.291*	0.216***	0.007
	(0.167)	*(0.083)*	*(0.162)*
	[0.066]	*[0.019]*	*[0.968]*
Treatment x Higher-SES	-0.299	-0.280**	0.105
	(0.275)	*(0.137)*	*(0.269)*
	[0.268]	*[0.041]*	*[0.684]*
Higher-SES	-0.086	0.116	-0.350
	(0.220)	*(0.109)*	*(0.223)*
Outcome in t_0	0.382***	0.490***	0.319***
	(0.089)	*(0.084)*	*(0.081)*
Randomization-pair fixed effects	Yes	Yes	Yes
Covariates	Yes	Yes	Yes
Observations	291	290	291
R^2	0.696	0.667	0.693
Treatment effect for Higher-SES	-0.008	-0.065	0.111
	(0.173)	*(0.091)*	*(0.169)*
SES gap	-0.077	-0.031	-0.059
Control-group mean	0.000	0.444	0.000

Notes: Table shows ITT effects of the mentoring program on labor-market orientation. Variables and indices are standardized by subtracting the control-group mean and dividing by the control-group standard deviation. Ordinary least squares estimates. *SES gap* is calculated as the coefficient on higher-SES background in a regression of the respective outcome on the higher-SES indicator in the control-group sample in the follow-up survey. *Control-group mean* indicates the mean of the respective outcome in the control-group sample in the follow-up survey. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in t_0 are included. Robust standard errors in parentheses. Randomization inference (RI) *p*-values in square brackets, obtained from RI with 1,000 permutations, assigning the treatment status randomly within randomization pairs. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix (for online publication only)

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

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

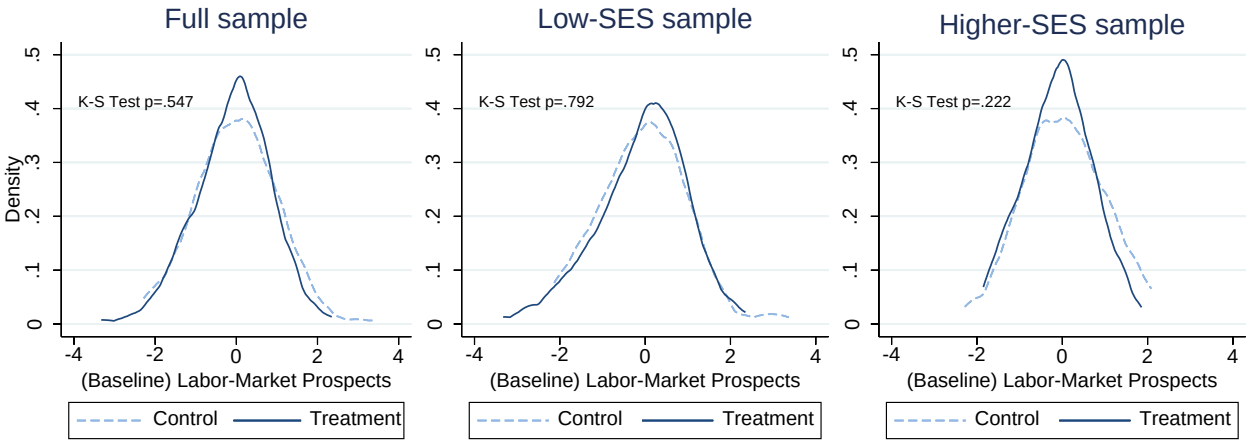
February 3, 2021

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

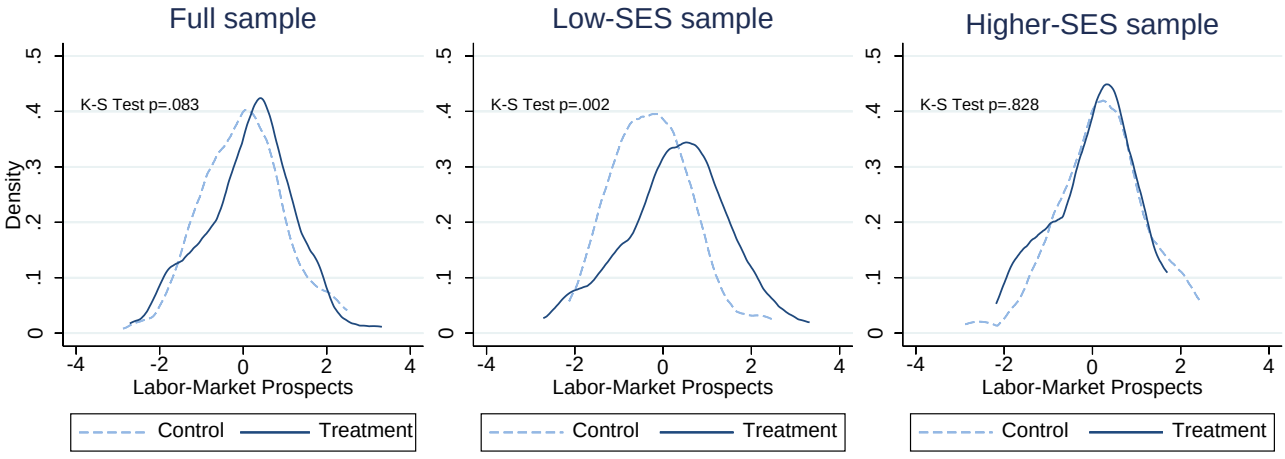
Appendix A: Additional Figures and Tables

Figure A1: Effect of the Mentoring Program on the Distribution of Labor-Market Prospects

Panel A: Distribution of labor-market prospects in baseline survey



Panel B: Distribution of labor-market prospects in follow-up survey



Notes: Panel A shows the entire distribution of the index of labor-market prospects for the treatment and control groups in the baseline (pre-treatment) survey. Panel B shows the unconditional treatment effect on the entire distribution of the index of labor-market prospects in the follow-up survey. Samples: all respondents (left), low-SES respondents (those with at most 25 books at home at baseline) (middle), higher-SES respondents (those with more than 25 books at home) (right). The probability density functions are computed with an Epanechnikov kernel with bandwidth h derived from the Silverman rule (Silverman (1986), pp. 47-48) with $h = 0.9An^{-1/5}$, where n is the number of observations and $A = min(standard\ deviation, interquartile\ range/1.349)$. K-S Test: p-values for a Kolmogorov-Smirnov test of the equality of distributions of labor-market prospects of treatment and control groups.

Table A1: Observations in the Baseline Survey by Mentoring Site and Cohort

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

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

Table A2: Observations in the Follow-up Survey by Mentoring Site and Cohort

Cohort	Site/city	School ID	Survey period		Survey sample		Administrative sample	
			Month	Year	Control	Treatment	Control	Treatment
1	Aachen[a]	1	11	2016	13	14	14	14
1	Aachen	1	11	2017	14	13	14	13
1	Berlin	1	11	2017	3	4	3	4
1	Berlin	2	5	2018	8	7	8	7
1	Berlin	3	5	2018	5	6	5	6
1	Cologne	1	11	2017	6	6	7	6
1	Cologne	2	11	2017	6	6	6	6
1	Cologne	3	11	2017	4	5	4	5
1	Duisburg[a,b]	1	9	2017	5	7	5	7
1	Essen[b]	1	12	2017	4	4	4	3
1	Hamburg	1	12	2017	5	4	4	3
1	Hamburg	2	2	2018	7	5	6	6
1	Hamburg	3	12	2017	1	2	2	2
1	Luebeck	1	12	2017	13	12	13	13
1	Luebeck	2	11	2017	8	8	8	8
1	Mannheim	1	10	2017	4	4	4	3
2	Berlin	1	12	2018	5	5	5	5
2	Berlin	3	5	2019	3	4	2	3
2	Cologne	1	11	2018	7	7	8	7
2	Cologne	2	11	2018	4	3	4	3
2	Hamburg	1	12	2018	2	2	2	1
2	Hamburg	2	12	2018	1	1	1	1
2	Hamburg	4	12	2018	4	4	4	5
2	Leipzig	1	6	2019	1	0	2	2
2	Leipzig	2	6	2019	0	0	0	0
2	Luebeck	1	12	2018	4	5	5	5
2	Lueneburg	2	12	2018	8	8	8	8
					145	146	148	146
					291		294	

Notes: Table shows dates and sample sizes of the follow-up survey for each site and cohort. Sample sizes refer to the randomized sample. *Survey sample*: number of observations in the survey. *Administrative sample*: number of observations in the administrative school records. [a] Pilot studies. [b] Duisburg and Essen belong to the same mentoring site.

Table A3: Comparison of Randomized and Non-Randomized Samples

	Sample			Difference		
	Total *N=442*	Randomized *N=308*	Non-rand. *N=134*	(2)-(3)	*p*-value	Obs.
	(3)	(1)	(2)	(4)	(5)	(6)
A. Outcome variables at baseline						
Overall index	-0.05	-0.04	-0.05	0.01	0.936	442
Components						
Math grade (administrative)	-0.03	0.01	-0.12	0.13	0.308	311
Math grade (admin.) missing d.	0.30	0.29	0.31	-0.01	0.773	442
Patience and social skills index	-0.07	-0.04	-0.16	0.12	0.279	442
Labor-market orientation index	0.02	-0.04	0.15	-0.20	0.064	441
B. Components of outcome variables at baseline						
Patience and social skills index						
Patience	-0.08	-0.01	-0.24	0.23	0.044	441
Social skills index	-0.03	-0.05	0.00	-0.05	0.643	442
Components						
Prosociality	0.06	0.01	0.17	-0.16	0.082	442
Trust	-0.03	-0.02	-0.06	0.03	0.754	438
Self-efficacy	-0.09	-0.08	-0.11	0.03	0.767	441
Labor-market orientation index						
Wants apprenticeship after school	0.40	0.37	0.45	-0.08	0.109	400
Knows future career	-0.04	-0.08	0.06	-0.14	0.156	439
C. Matching and balancing variables for randomization at baseline						
Male	0.43	0.44	0.43	0.00	0.965	442
Age	14.00	13.98	14.04	-0.06	0.504	442
Migrant	0.57	0.58	0.57	0.01	0.834	442
Books at home	1.72	1.70	1.76	-0.06	0.409	442
Math grade (survey)	1.74	1.72	1.77	-0.05	0.580	361
Math grade (survey) missing d.	0.18	0.15	0.25	-0.10	0.019	442
German grade (survey)	1.74	1.72	1.79	-0.07	0.368	359
German grade (survey) missing d.	0.19	0.16	0.25	-0.08	0.051	442
English grade (survey)	1.80	1.81	1.76	0.05	0.516	359
English grade (survey) missing d.	0.19	0.16	0.25	-0.08	0.051	442
Received paid private teaching	0.20	0.19	0.20	0.00	0.987	441
Parental homework support	2.75	2.76	2.74	0.02	0.868	441
Big-5: Conscientiousness	3.28	3.31	3.22	0.09	0.289	442
Big-5: Neuroticism	2.92	2.94	2.86	0.08	0.372	442
D. Further control variables at baseline						
Big-5: Openness	3.46	3.46	3.46	0.00	0.992	442
Big-5: Extraversion	3.35	3.33	3.40	-0.07	0.460	441
Big-5: Agreeableness	3.47	3.48	3.43	0.05	0.532	441
Higher SES (>25 books at home)	0.55	0.53	0.58	-0.05	0.304	442

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

Table A4: Comparison of Sample to German Student Population in PISA

	All		Low-SES			Higher-SES		
	Mean	Δ PISA total	Mean	Δ PISA total	Δ PISA low-SES	Mean	Δ PISA total	Δ PISA higher-SES
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Low-SES (<=25 books at home)	0.47	0.24***	1.00	0.77***	—	0.00	-0.23***	—
Books at home								
0-10 books	0.24	0.15***	0.52	0.42***	0.09**	—	-0.10***	—
11-25 books	0.23	0.10***	0.48	0.35***	-0.09**	—	-0.13***	—
26-100 books	0.25	-0.03	—	-0.28***	—	0.48	0.20***	0.11***
101-200 books	0.11	-0.10***	—	-0.20***	—	0.20	0.00	-0.06*
201-500 books	0.09	-0.09***	—	-0.18***	—	0.17	-0.01	-0.06*
More than 500 books	0.08	-0.03	—	-0.11***	—	0.15	0.04	0.01
Student demographics								
Male	0.44	-0.06**	0.47	-0.03	-0.09**	0.40	-0.09**	-0.07*
Age	13.98	-1.85***	14.25	-1.58***	-1.58***	13.74	-2.09***	-2.09***
Migrant	0.58	0.30***	0.72	0.44***	0.30***	0.45	0.17***	0.21***
University degree, father								
Yes	0.22	-0.01	0.15	-0.08**	0.05**	0.28	0.05	0.01
Missing	0.40	0.26***	0.41	0.28***	0.17***	0.39	0.25***	0.28***
University degree, mother								
Yes	0.23	0.06***	0.12	-0.05*	0.05**	0.33	0.15***	0.12***
Missing	0.32	0.22***	0.31	0.21***	0.11***	0.34	0.24***	0.26***
Employment, father								
Full-time	0.70	-0.11***	0.68	-0.13***	-0.06*	0.72	-0.08**	-0.10***
Part-time	0.06	-0.01	0.03	-0.04*	-0.05**	0.09	0.02	0.02
Not employed, not searching	0.03	-0.01	0.04	0.00	0.00	0.02	-0.02	-0.02
Unemployed	0.03	0.00	0.05	0.02*	0.00	0.01	-0.02	-0.01
Missing	0.19	0.12***	0.21	0.14***	0.11***	0.17	0.10***	0.11***
Employment, mother								
Full-time	0.39	0.07***	0.34	0.02	0.02	0.44	0.12***	0.12***
Part-time	0.25	-0.17***	0.27	-0.15***	-0.09**	0.23	-0.19***	-0.21***
Not employed, not searching	0.21	0.03	0.23	0.05	0.01	0.19	0.01	0.02
Unemployed	0.06	0.02*	0.07	0.03*	0.02	0.06	0.02	0.02
Missing	0.09	0.05***	0.09	0.05***	0.03	0.09	0.05***	0.05***

Notes: Table shows group means for our baseline sample (*N*=308) and differences to the PISA 2012 sample. Columns 1-2: full sample; columns 3-5: low-SES sample (respondents with at most 25 books at home); columns 6-8: higher-SES sample (respondents with more than 25 books at home). Significance of the difference is tested by a two-side *t*-test on the mean. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A5: Variable Definitions and Wording of Questionnaire Items

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

(continued on next page)

Table A5 (continued)

	Wording (English translation) (1)	Wording (German original) (2)	Answer categories (3)
Labor-market orientation index	*Equally weighted average of two components: wants apprenticeship after school; knows future career*		
Wants apprenticeship after school	Answer "Apprenticeship" to "What would you like to do after finishing school?"	Was möchtest du nach deinem gewünschten Schulabschluss machen? Ausbildung.	Apprenticeship; university; directly entering a job; something else; don't know yet
Knows future career	Agreement to item: I already know exactly which occupation I want to work in later in life.	Ich weiß schon genau, was ich später mal beruflich machen will.	4-point scale from "do not agree at all" to "agree"
Covariates (baseline survey)			
Higher-SES (>25 books at home)	Answer more than 25 to "Approximately how many books are there in your home?"	Wie viele Bücher gibt es bei dir zuhause ungefähr? Antworten: genug, um mehrere Regalbretter zu füllen (26 bis 100 Bücher), genug, um ein kleines Regal zu füllen (101 bis 200 Bücher), genug, um ein großes Regal zu füllen (201 bis 500 Bücher), genug, um eine Regalwand zu füllen (mehr als 500 Bücher).	Books: 0-10; 11-25; 26-100; 101-200; 201-500; more than 500
Male	Answer "male" to "Are you male or female?"	Bist du männlich oder weiblich?	Male; female
Age	Based on "When were you born?"	Wann bist du geboren?	Day, month, and year of birth
Migrant	Adolescent or at least one parent not born in Germany.	In welchem Land bist du geboren? In welchem Land ist deine Mutter geboren? In welchem Land ist dein Vater geboren?	Germany; other country (name)
Received paid private teaching	Answer "Yes" to "Did you get paid private teaching in the last semester of school?"	Hast du im letzten Schulhalbjahr bezahlten Nachhilfeunterricht bekommen?	Yes; no
Parental homework support	Do your parents (mother and/or father) support you with your homework and learning for school?	Unterstützen dich deine Eltern (Mutter und/oder Vater) bei den Hausaufgaben und beim Lernen für die Schule?	4-point scale: not at all; rather little; rather strong; very strong

(continued on next page)

Table A5 (continued)

	Wording (English translation) (1)	Wording (German original) (2)	Answer categories (3)
Big-5 personality traits	*Personality inventory according to 10-Item Big-5 Inventory (Rammstedt (2007); Rammstedt and John (2007))*		
Conscientiousness	Agreement to two items: I am someone who tends to be lazy (reversed); I am someone who does a thorough job.	Ich bin bequem, neige zur Faulheit; Ich erledige Aufgaben gründlich.	5-point scales from "does not apply at all" to "applies completely"
Neuroticism	Agreement to two items: I am someone who is relaxed, handles stress well (reversed); I am someone who gets nervous easily.	Ich bin entspannt, lasse mich durch Stress nicht aus der Ruhe bringen; Ich werde leicht nervös und unsicher.	5-point scales from "does not apply at all" to "applies completely"
Openness	Agreement to two items: I am someone who has few artistic interests (reversed); I am someone who has a vivid imagination/fantasy.	Ich habe nur wenig künstlerisches Interesse; Ich habe eine aktive Vorstellungskraft, bin fantasievoll.	5-point scales from "does not apply at all" to "applies completely"
Extraversion	Agreement to two items: I am someone who is reserved (reversed); I am someone who is outgoing, sociable.	Ich bin eher zurückhaltend, reserviert; Ich gehe aus mir heraus, bin gesellig.	5-point scales from "does not apply at all" to "applies completely"
Agreeableness	Agreement to two items: I am someone who is generally trusting; I am someone who tends to find fault with others (reversed).	Ich schenke anderen leicht Vertrauen, glaube an das Gute im Menschen; Ich neige dazu, andere zu kritisieren.	5-point scales from "does not apply at all" to "applies completely"

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

Table A6: Balancing in Subsamples

	Low-SES sample			Higher-SES sample		
	Control Mean	Treatment Mean	Difference *p*-value	Control Mean	Treatment Mean	Difference *p*-value
	(1)	(2)	(3)	(4)	(5)	(6)
A. Outcome variables at baseline						
Overall index	-0.05	-0.08	0.872	0.04	-0.10	0.316
Components						
Math grade (administrative)	-0.19	0.05	0.198	0.20	-0.02	0.227
Math grade (admin.) missing d.	0.22	0.22	0.965	0.34	0.37	0.702
Patience and social skills index	-0.07	-0.15	0.648	0.06	0.00	0.694
Labor-market orientation index	0.14	-0.02	0.331	-0.13	-0.15	0.862
B. Components of outcome variables at baseline						
Patience and social skills index						
Patience	0.04	-0.07	0.530	-0.04	0.03	0.649
Social skills index	-0.15	-0.15	0.959	0.13	-0.03	0.251
Components						
Prosociality	-0.07	-0.04	0.841	0.06	0.06	0.978
Trust	-0.12	-0.13	0.917	0.10	0.03	0.607
Self-efficacy	-0.12	-0.15	0.828	0.10	-0.16	0.080
Labor-market orientation index						
Wants apprenticeship after school	0.44	0.44	0.920	0.29	0.31	0.771
Knows future career	0.08	-0.16	0.156	-0.07	-0.16	0.567
C. Matching and balancing variables for randomization at baseline						
Male	0.45	0.49	0.684	0.41	0.40	0.801
Age	14.23	14.26	0.839	13.77	13.70	0.633
Migrant	0.71	0.72	0.896	0.48	0.43	0.580
Books at home	1.00	1.00		2.38	2.26	0.105
Math grade (survey)	1.79	1.77	0.910	1.63	1.69	0.614
Math grade (survey) missing d.	0.04	0.08	0.296	0.23	0.23	0.966
German grade (survey)	1.76	1.80	0.672	1.70	1.61	0.356
German grade (survey) missing d.	0.04	0.08	0.296	0.26	0.25	0.893
English grade (survey)	1.86	2.03	0.149	1.73	1.63	0.376
English grade (survey) missing d.	0.05	0.10	0.339	0.24	0.23	0.890
Received paid private teaching	0.18	0.18	0.969	0.18	0.23	0.421
Parental homework support	2.58	2.56	0.871	3.00	2.84	0.246
Big-5: Conscientiousness	3.47	3.26	0.117	3.25	3.27	0.861
Big-5: Neuroticism	2.86	2.94	0.528	2.95	3.01	0.598
D. Further control variables at baseline						
Big-5: Openness	3.30	3.27	0.856	3.52	3.73	0.125
Big-5: Extraversion	3.26	3.33	0.609	3.35	3.38	0.834
Big-5: Agreeableness	3.53	3.49	0.746	3.48	3.44	0.824
Higher SES (>25 books at home)	0.00	0.00		1.00	1.00	

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

Table A7: Overall Effect of the Mentoring Program: Average Effects

	ITT				TOT
	(1)	(2)	(3)	(4)	(5)
Treatment	0.149	0.199**	0.193**	0.153*	0.177*
	(0.119)	*(0.100)*	*(0.091)*	*(0.089)*	*(0.103)*
	[0.140]	*[0.030]*	*[0.035]*	*[0.118]*	–
Outcome in t_0		0.585***	0.480***	0.441***	0.429***
		(0.051)	*(0.065)*	*(0.082)*	*(0.083)*
Randomization-pair fixed effects	No	No	Yes	Yes	Yes
Covariates	No	No	No	Yes	Yes
Observations	304	304	304	304	304
R^2	0.005	0.295	0.703	0.730	0.731
Kleibergen-Paap *F* statistic					858.65

Notes: Table shows intent-to-treat effects (ITT) and treatment-on-treated (TOT) effects of the mentoring program on the index of labor-market prospects. Index is an equally weighted average of *z*-scores of its components: administrative math grade (reversed), patience and social skills index, and labor-market orientation index (see Kling, Liebman, and Katz (2007)). Calculation of each *z*-score subtracts the score's control group mean and divides by the control group standard deviation. Ordinary least squares estimates in columns 1-4, two-stage least squares estimates in column 5. In the TOT estimation in column 5, *Treatment* indicates program take-up (one if mentor and mentee have met at least once, zero otherwise), which is instrumented by the random treatment assignment. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in t_0 are included. Robust standard errors in parentheses. Randomization inference (RI) *p*-values in square brackets, obtained from RI with 1,000 permutations, assigning the treatment status randomly within randomization pairs. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A8: Effect of the Mentoring Program on Main Outcomes

	Outcome index	Math grade	Patience and social skills index	Labor-market orientation index
	(1)	(2)	(3)	(4)
Treatment	0.556***	0.294**	0.439***	0.291*
	(0.143)	*(0.142)*	*(0.152)*	*(0.167)*
Treatment x Higher-SES	-0.748***	-0.467**	-0.587**	-0.299
	(0.220)	*(0.230)*	*(0.255)*	*(0.275)*
Higher-SES	0.182	0.283	0.164	-0.086
	(0.195)	*(0.196)*	*(0.227)*	*(0.220)*
Outcome in t_0	0.459***	0.488***	0.261**	0.382***
	(0.078)	*(0.100)*	*(0.103)*	*(0.089)*
Male	-0.215	-0.116	-0.414	0.248
	(0.217)	*(0.216)*	*(0.291)*	*(0.287)*
Age	-0.019	0.008	0.051	-0.063
	(0.096)	*(0.091)*	*(0.105)*	*(0.093)*
Migrant	-0.177	-0.135	0.001	-0.121
	(0.160)	*(0.136)*	*(0.159)*	*(0.206)*
Received paid private teaching	0.226	-0.158	0.155	0.252
	(0.166)	*(0.142)*	*(0.194)*	*(0.193)*
Parental homework support	0.011	-0.085	-0.027	0.162*
	(0.072)	*(0.063)*	*(0.082)*	*(0.086)*
Big-5: openness	0.206***	0.004	0.077	0.267***
	(0.077)	*(0.065)*	*(0.090)*	*(0.091)*
Big-5: conscientiousness	0.113	0.110	0.105	0.121
	(0.108)	*(0.076)*	*(0.116)*	*(0.102)*
Big-5: extraversion	0.092	0.003	0.081	0.117
	(0.091)	*(0.082)*	*(0.104)*	*(0.105)*
Big-5: agreeableness	-0.117	-0.067	0.115	-0.169*
	(0.085)	*(0.076)*	*(0.098)*	*(0.097)*
Big-5: neuroticism	-0.065	-0.026	-0.167	-0.008
	(0.085)	*(0.084)*	*(0.108)*	*(0.109)*
Constant	-0.035	1.029	-1.561	-0.455
	(1.514)	*(1.312)*	*(1.648)*	*(1.600)*
Randomization-pair fixed effects	Yes	Yes	Yes	Yes
Observations	304	294	291	291
R^2	0.753	0.775	0.695	0.696

Notes: Table shows ITT effects of the mentoring program on the outcome indicated in the column header. Covariates are from the baseline survey. Dummies for missing values in t_0 are included. Robust standard errors in parentheses. Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A9: Overall Effect of the Mentoring Program: Splitting the Sample by SES Status

	SES		SES pair	
	Low-SES	Higher-SES	Low-SES	Higher-SES
	(1)	(2)	(3)	(4)
A. Without randomization-pair fixed effects				
Treatment	0.613***	-0.164	0.719***	-0.190
	(0.146)	*(0.132)*	*(0.168)*	*(0.153)*
Outcome in t_0	0.542***	0.454***	0.612***	0.477***
	(0.074)	*(0.094)*	*(0.089)*	*(0.116)*
Randomization-pair fixed effects	No	No	No	No
Covariates	Yes	Yes	Yes	Yes
Observations	142	162	84	102
R^2	0.498	0.315	0.660	0.407
B. With randomization-pair fixed effects				
Treatment	0.644***	-0.224	0.641***	-0.195
	(0.213)	*(0.184)*	*(0.170)*	*(0.156)*
Outcome in t_0	0.642***	0.406**	0.639***	0.373*
	(0.149)	*(0.194)*	*(0.131)*	*(0.208)*
Randomization-pair fixed effects	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes
Observations	142	162	84	102
R^2	0.916	0.827	0.882	0.716

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

Table A10: TOT Effect of the Mentoring Program on Main Outcomes

	Outcome index	Math grade	Patience and social skills index	Labor-market orientation index
	(1)	(2)	(3)	(4)
Treatment	0.684***	0.366**	0.540***	0.355*
	(0.170)	*(0.174)*	*(0.183)*	*(0.205)*
Treatment x Higher-SES	-0.895***	-0.556**	-0.703**	-0.363
	(0.251)	*(0.268)*	*(0.293)*	*(0.322)*
Higher-SES	0.201	0.299	0.175	-0.081
	(0.196)	*(0.199)*	*(0.228)*	*(0.223)*
Outcome in t_0	0.439***	0.473***	0.264**	0.375***
	(0.078)	*(0.103)*	*(0.103)*	*(0.089)*
Randomization-pair fixed effects	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes
Observations	304	294	291	291
R^2	0.754	0.777	0.695	0.696
Kleibergen-Paap *F* statistic	141.50	109.16	119.02	126.63
Treatment effect for Higher-SES	-0.211	-0.190	-0.164	-0.007
	(0.151)	*(0.159)*	*(0.195)*	*(0.195)*

Notes: Table shows TOT effects of the mentoring program on the outcome indicated in the column header. *Treatment* indicates program take-up (one if mentor and mentee have met at least once, zero otherwise), which is instrumented by the random treatment assignment. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in t_0 are included. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A11: Effect of the Mentoring Program on Satisfaction Outcomes

Satisfaction with:	Life	Performance in math	Performance in German	Performance in English	Own money and wealth	Family	Friends	School
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.224**	0.122	-0.037	0.122	0.157**	0.031	-0.015	0.061
	(0.093)	*(0.109)*	*(0.110)*	*(0.083)*	*(0.074)*	*(0.085)*	*(0.097)*	*(0.099)*
Treatment x Higher-SES	-0.272**	-0.119	0.127	-0.070	-0.215**	-0.111	-0.033	-0.074
	(0.128)	*(0.152)*	*(0.167)*	*(0.135)*	*(0.102)*	*(0.120)*	*(0.137)*	*(0.157)*
Higher-SES	0.072	-0.005	-0.034	0.211*	0.138	-0.084	0.044	-0.010
	(0.107)	*(0.116)*	*(0.129)*	*(0.115)*	*(0.086)*	*(0.098)*	*(0.103)*	*(0.122)*
Outcome in t_0	0.147	0.231***	0.255**	0.591***	0.199**	0.198**	0.219*	0.349***
	(0.098)	*(0.083)*	*(0.101)*	*(0.077)*	*(0.086)*	*(0.097)*	*(0.119)*	*(0.098)*
Randomization-pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	291	291	291	290	291	290	291	291
R^2	0.635	0.635	0.540	0.691	0.639	0.617	0.530	0.587
Treatment effect for Higher-SES	-0.048	0.003	0.090	0.051	-0.058	-0.079	-0.048	-0.012
	(0.071)	*(0.081)*	*(0.102)*	*(0.091)*	*(0.054)*	*(0.066)*	*(0.074)*	*(0.098)*

Notes: Table shows ITT effects of the mentoring program on satisfaction domains indicated in the column header. Satisfaction in each domain is report on a 5-point Likert scale, ranging from "totally dissatisfied" to "totally satisfied". Dependent variables are dummies, which are one if the individual reports that they are "somewhat satisfied" or "totally satisfied" and zero otherwise. *Life:* How satisfied are you currently, all in all, with your life? *Performance:* How satisfied are you with your performance in […]. *Own money and wealth:* How satisfied are you with your belongings? Think about money and things that you own. *Family:* How satisfied are you with your family? *Friends:* How satisfied are you with your circle of friends? *School:* How satisfied are you with your situation in school? Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in t_0 are included. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A12: Effect of the Mentoring Program on the Three Components of the Index of Labor-Market Prospects: Correction for Multiple Hypotheses Testing

	Math grade	Patience and social skills	Labor-market orientation
	(1)	(2)	(3)
Treatment	0.294	0.439	0.291
Standard *p*-value	*0.039*	*0.004*	*0.084*
List-Shaikh-Xu *p*-value	*0.059*	*0.021*	*0.087*
Westfall-Young *p*-value	*0.069*	*0.008*	*0.070*
Treatment x Higher-SES	-0.467	-0.587	-0.299
Standard *p*-value	*0.044*	*0.023*	*0.279*
List-Shaikh-Xu *p*-value	*0.081*	*0.067*	*0.287*
Westfall-Young *p*-value	*0.081*	*0.074*	*0.278*
Higher-SES dummy	Yes	Yes	Yes
Outcome in t_0	Yes	Yes	Yes
Randomization-pair fixed effects	Yes	Yes	Yes
Covariates	Yes	Yes	Yes
Observations	294	291	291

Notes: Table shows ITT effects of the mentoring program on administrative math grade (reversed), patience and social skills index, and labor-market orientation index. The three columns replicate the specifications in the first column of Tables 3, 4, and 5, respectively. In addition to the standard *p*-values based on robust standard errors, the table reports *p*-values robust to multiple hypothesis testing (family-wise error rates) using the bootstrap resampling techniques by List, Shaikh, and Xu (2019) and Westfall and Young (1993), respectively. Bootstraps are adjusted to account for the pair structure in the data, i.e., the sample drawn during each replication is a bootstrap sample of pairs. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in t_0 are included.

Table A13: Effect of the Mentoring Program on Measures of Social Capital

	Volunteer	Friends		Meet friends			
		No friends	Number of friends	Never	Meet often	Number of meetings	# meetings conditional on # friends
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.026	-0.024	-0.139	-0.021	0.042	0.164	0.290
	(0.081)	*(0.031)*	*(0.152)*	*(0.028)*	*(0.098)*	*(0.226)*	*(0.228)*
Treatment x Higher-SES	-0.009	0.022	-0.018	0.025	-0.079	-0.227	-0.301
	(0.115)	*(0.049)*	*(0.249)*	*(0.044)*	*(0.149)*	*(0.306)*	*(0.305)*
Higher-SES	0.070	-0.063	0.069	-0.068	0.102	0.299	0.333
	(0.095)	*(0.042)*	*(0.247)*	*(0.043)*	*(0.117)*	*(0.276)*	*(0.267)*
Outcome in t_0	0.550***	0.170*	0.241**	0.220**	0.452***	0.350***	0.317***
	(0.076)	*(0.098)*	*(0.099)*	*(0.110)*	*(0.089)*	*(0.110)*	*(0.110)*
Friends in t_1							0.023*
							(0.013)
Friends in t_0							0.015
							(0.015)
Randomization-pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	290	273	284	287	287	287	287
R^2	0.763	0.619	0.571	0.560	0.614	0.602	0.637
Treatment effect for Higher-SES	0.017	-0.002	-0.157	0.004	-0.038	-0.063	-0.010
	(0.068)	*(0.037)*	*(0.165)*	*(0.035)*	*(0.091)*	*(0.172)*	*(0.173)*

Notes: Table shows ITT effects of the mentoring program on social capital domains indicated in the column header. *Volunteer:* Are you volunteering at least once a week in a club or association (e.g., sports clubs, youth clubs, voluntary fire brigade, supporters club, political parties, musical and artistic groups, etc.)? *Friends:* How many friends do you regularly meet in your private time, i.e., outside of school time. *No friends:* Dummy variable that is one if the individual reports to meet with zero friends and zero otherwise. *Number of friends:* Standardized number of friends. *Meet friends:* How many times do you meet with friends in your private time, i.e., outside of school time. *Never:* Dummy variable that is one if the individual reports to have no regular meetings and zero otherwise. *Meet often:* Dummy variable that is one if the individual reports to have meetings for at least two days in a regular week and zero otherwise. *Number of meetings:* Standardized number of meeting days in a regular week. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in t_0 are included. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A14: Effect of the Mentoring Program on Measures of Social Capital in School

	All activities	High-stakes activities	Low-stakes activities
	(1)	(2)	(3)
Treatment	0.023	0.040	-0.003
	(0.201)	*(0.205)*	*(0.210)*
Treatment x Higher-SES	-0.343	-0.157	-0.330
	(0.309)	*(0.311)*	*(0.321)*
Higher-SES	0.115	0.050	0.114
	(0.241)	*(0.226)*	*(0.257)*
Randomization-pair fixed effects	Yes	Yes	Yes
Covariates	Yes	Yes	Yes
Observations	290	290	290
R^2	0.634	0.641	0.562
Treatment effect for Higher-SES	-0.320*	-0.118	-0.334*
	(0.185)	*(0.183)*	*(0.197)*

Notes: Table shows ITT effects of the mentoring program on social capital in school. Dependent variable in column 1 is an average index of the following school activities (represented by a dummy variable that is one if true and zero otherwise): acting as class representative, working as peer mediator, acting as school representative, working for the school magazine, volunteering as school nurse, participating in the school music ensemble, participating in the school theater group, and participating in other school activity. Dependent variable in column 2 includes only more high-stakes activities: acting as class representative, working as peer mediator, and acting as school representative. Dependent variable in column 3 collects the remaining activities. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in t_0 are included. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A15: Effect of the Mentoring Program on Plans after School

	Apprenticeship	University	Don't know	Direct job	Other
	(1)	(2)	(3)	(4)	(5)
Treatment	0.216***	-0.111	-0.115	0.025	0.003
	(0.083)	*(0.087)*	*(0.086)*	*(0.042)*	*(0.048)*
Treatment x Higher-SES	-0.280**	0.236*	0.081	-0.065	-0.009
	(0.137)	*(0.131)*	*(0.140)*	*(0.058)*	*(0.066)*
Higher-SES	0.116	-0.084	-0.030	0.014	-0.003
	(0.109)	*(0.109)*	*(0.109)*	*(0.044)*	*(0.051)*
Outcome in t_0	0.490***	0.469***	0.318***	-0.038	0.172
	(0.084)	*(0.094)*	*(0.095)*	*(0.033)*	*(0.186)*
Randomization-pair fixed effects	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes
Observations	290	290	290	290	290
R^2	0.667	0.646	0.587	0.578	0.537
Treatment effect for Higher-SES	-0.065	0.124	-0.034	-0.040	-0.006
	(0.091)	*(0.084)*	*(0.093)*	*(0.029)*	*(0.037)*

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

Table A16: Effect of the Mentoring Program by Gender

	(1)	(2)	(3)
Treatment	0.556***	0.212	0.569***
	(0.143)	*(0.171)*	*(0.192)*
Treatment x Higher-SES	-0.748***		-0.746***
	(0.220)		*(0.223)*
Treatment x Female		-0.104	-0.025
		(0.215)	*(0.211)*
Higher-SES	0.182		0.182
	(0.195)		*(0.196)*
Female	0.215	0.257	0.227
	(0.217)	*(0.241)*	*(0.240)*
Outcome in t_0	0.459***	0.440***	0.459***
	(0.078)	*(0.083)*	*(0.078)*
Randomization-pair fixed effects	Yes	Yes	Yes
Covariates	Yes	Yes	Yes
Observations	304	304	304
R^2	0.753	0.731	0.753
Treatment effect for Higher-SES	-0.192		-0.177
	(0.137)		*(0.210)*
Treatment effect for Females		0.108	
		(0.108)	
Treatment effect for Females, low-SES			0.544***
			(0.163)

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

Table A17: Overall Effect of the Mentoring Program: Using Three SES Categories

	ITT				TOT
	(1)	(2)	(3)	(4)	(5)
Treatment	0.549***	0.573***	0.538***	0.546***	0.673***
	(0.181)	*(0.144)*	*(0.137)*	*(0.143)*	*(0.171)*
Treatment x Medium-SES	-0.820***	-0.640***	-0.550**	-0.717***	-0.856***
	(0.263)	*(0.224)*	*(0.252)*	*(0.264)*	*(0.299)*
Treatment x High-SES	-0.600*	-0.812***	-0.871***	-0.868***	-1.026***
	(0.316)	*(0.258)*	*(0.318)*	*(0.323)*	*(0.347)*
Medium-SES	0.513***	0.381**	0.062	0.222	0.237
	(0.180)	*(0.166)*	*(0.207)*	*(0.223)*	*(0.222)*
High-SES	0.415*	0.510***	0.137	0.043	0.074
	(0.215)	*(0.181)*	*(0.253)*	*(0.263)*	*(0.263)*
Outcome in t_0		0.584***	0.506***	0.454***	0.434***
		(0.053)	*(0.064)*	*(0.077)*	*(0.077)*
Randomization-pair fixed effects	No	No	Yes	Yes	Yes
Covariates	No	No	No	Yes	Yes
Observations	304	304	304	304	304
R^2	0.041	0.326	0.725	0.756	0.757
Kleibergen-Paap *F* statistic					106.21
Treatment effect for Medium-SES	-0.271	-0.067	-0.013	-0.171	-0.183
	(0.190)	*(0.173)*	*(0.189)*	*(0.189)*	*(0.212)*
Treatment effect for High-SES	-0.051	-0.239	-0.333	-0.321	-0.353
	(0.259)	*(0.214)*	*(0.284)*	*(0.286)*	*(0.294)*

Notes: Table shows intent-to-treat effects (ITT) and treatment-on-treated (TOT) effects of the mentoring program on the index of labor-market prospects. Ordinary least squares estimates in columns 1-4, two-stage least squares estimates in column 5. In the TOT estimation in column 5, *Treatment* indicates program take-up (one if mentor and mentee have met at least once, zero otherwise), which is instrumented by the random treatment assignment. Medium-SES (High-SES) are characterized by reporting 26 to 200 (more than 200) books at home in the baseline survey. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in t_0 are included. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A18: Effect of the Mentoring Program: Accounting for Possible Spillover Effects

	Baseline	School	Class	
		Absolute	Absolute	Share
	(1)	(2)	(3)	(4)
Treatment	0.573***	0.569***	0.570***	0.572***
	(0.144)	*(0.159)*	*(0.147)*	*(0.147)*
Treatment x Treated pupils		-0.015	-0.041	0.058
		(0.187)	*(0.161)*	*(0.160)*
Treatment x Higher-SES	-0.701***	-0.687***	-0.719***	-0.714***
	(0.197)	*(0.214)*	*(0.200)*	*(0.200)*
Treatment x Higher-SES x Treated pupils		-0.026	0.051	-0.125
		(0.229)	*(0.206)*	*(0.212)*
Higher-SES	0.431***	0.414***	0.430***	0.432***
	(0.140)	*(0.153)*	*(0.143)*	*(0.143)*
Treated pupils		-0.032	0.015	0.011
		(0.136)	*(0.097)*	*(0.093)*
Treated pupils x Higher-SES		0.130	0.092	0.121
		(0.165)	*(0.135)*	*(0.137)*
Outcome in t_0	0.580***	0.577***	0.579***	0.583***
	(0.053)	*(0.053)*	*(0.054)*	*(0.054)*
Randomization-pair fixed effects	No	No	No	No
Covariates	No	No	No	No
Observations	304	304	304	304
R^2	0.325	0.329	0.332	0.332
Treatment effect for Low-SES	0.573***	0.554*	0.529**	0.629***
	(0.144)	*(0.294)*	*(0.236)*	*(0.237)*
Treatment effect for Higher-SES	-0.128	-0.159	-0.139	-0.210
	(0.135)	*(0.158)*	*(0.187)*	*(0.196)*

Notes: Table shows ITT effects of the mentoring program on the index of labor-market prospects. Treatment effect is interacted with *treated pupils*. Variable represents the number of participants in the mentoring program in the same school-cohort (column 2), in the same class (column 3), and in the same class relative to the total class size (column 4). Treated pupils is standardized by subtracting the mean and dividing by the standard deviation of the estimation sample. Dummies for missing values in t_0 are included. Robust standard errors in parentheses. Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A19: Possible Spillover Effects of the Mentoring Program on the Control Group

	School		Class			
	Absolute		Absolute		Share	
	(1)	(2)	(3)	(4)	(5)	(6)
Treated pupils	0.107	0.027	0.045	-0.000	0.048	-0.014
	(0.084)	*(0.150)*	*(0.070)*	*(0.098)*	*(0.072)*	*(0.096)*
Treated pupils x Higher-SES		0.083		0.071		0.098
		(0.171)		*(0.141)*		*(0.147)*
Higher-SES		0.524***		0.538***		0.541***
		(0.163)		*(0.156)*		*(0.157)*
Outcome in t_0	0.368***	0.328***	0.377***	0.330***	0.379***	0.331***
	(0.087)	*(0.086)*	*(0.086)*	*(0.084)*	*(0.085)*	*(0.083)*
Covariates	No	No	No	No	No	No
Observations	152	152	152	152	152	152
R^2	0.321	0.374	0.314	0.372	0.315	0.372

Notes: Table shows effects of the mentoring program on the index of labor-market prospects of the control group. *Treated pupils* represents the number of participants in the mentoring program in the same school-cohort (column 1 and 2), in the same class (column 3 and 4), and in the same class relative to the total class size (column 5 and 6). Treated pupils is standardized by subtracting the mean and dividing by the standard deviation of the estimation sample. Dummies for missing values in t_0 are included. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A20: Effect of the Mentoring Program on Extended Labor-Market Orientation Index

	Index	Wants apprenticeship after school	Knows future career	Apply for apprenticeship	Participation in labor-market orientation event	Agency and school important for job information
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.493***	0.216***	0.007	0.037	0.112	0.274
	(0.182)	*(0.083)*	*(0.162)*	*(0.102)*	*(0.096)*	*(0.219)*
Treatment x Higher-SES	-0.575**	-0.280**	0.105	-0.090	-0.172	-0.298
	(0.258)	*(0.137)*	*(0.269)*	*(0.147)*	*(0.139)*	*(0.299)*
Higher-SES	0.200	0.116	-0.350	0.032	0.037	0.193
	(0.220)	*(0.109)*	*(0.223)*	*(0.121)*	*(0.119)*	*(0.269)*
Outcome in t_0	0.526***	0.490***	0.319***	0.323***	—	0.483***
	(0.092)	*(0.084)*	*(0.081)*	*(0.085)*		*(0.096)*
Randomization-pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Observations	291	290	291	289	291	291
R^2	0.720	0.667	0.693	0.609	0.567	0.625
Treatment effect for Higher-SES	-0.083	-0.065	0.111	-0.053	-0.060	-0.024
	(0.151)	*(0.091)*	*(0.169)*	*(0.085)*	*(0.084)*	*(0.175)*

Notes: Table shows ITT effects of the mentoring program on labor-market orientation. Columns 2 and 3 are repeated from columns 2 and 3 of Table 5. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in t_0 are included. Robust standard errors in parentheses. Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A21: Overall Effect of the Mentoring Program: Using Extended Labor-Market Orientation Index

	ITT				TOT
	(1)	(2)	(3)	(4)	(5)
Treatment	0.421**	0.502***	0.556***	0.592***	0.726***
	(0.178)	*(0.142)*	*(0.126)*	*(0.136)*	*(0.167)*
Treatment x Higher-SES	-0.603**	-0.556***	-0.687***	-0.795***	-0.950***
	(0.235)	*(0.197)*	*(0.192)*	*(0.204)*	*(0.237)*
Higher-SES	0.398**	0.415***	0.197	0.286	0.303
	(0.159)	*(0.141)*	*(0.185)*	*(0.189)*	*(0.192)*
Outcome in t_0		0.571***	0.534***	0.495***	0.478***
		(0.054)	*(0.070)*	*(0.082)*	*(0.080)*
Randomization-pair fixed effects	No	No	Yes	Yes	Yes
Covariates	No	No	No	Yes	Yes
Observations	304	304	304	304	304
R^2	0.026	0.307	0.741	0.766	0.760
Kleibergen-Paap *F* statistic					148.65
Treatment effect for Higher-SES	-0.183	-0.054	-0.131	-0.202	-0.224
	(0.153)	*(0.136)*	*(0.129)*	*(0.130)*	*(0.144)*

Notes: Table shows intent-to-treat effects (ITT) and treatment-on-treated (TOT) effects of the mentoring program on the index of labor-market prospects, using an extended definition of labor-market orientation. In addition to *wants apprenticeship after school* and *knows future career*, the labor-market orientation index comprises *apply for apprenticeship*, *participation in labor-market orientation event*, and *agency and school important for job information* (see Table A20). Index is an equally weighted average of *z*-scores of its components: administrative math grade (reversed), patience and social skills index, and extended labor-market orientation index (see Kling, Liebman, and Katz (2007)). Calculation of each *z*-score subtracts the score's control group mean and divides by the control group standard deviation. Ordinary least squares estimates in columns 1-4, two-stage least squares estimates in column 5. In the TOT estimation in column 5, *Treatment* indicates program take-up (one if mentor and mentee have met at least once, zero otherwise), which is instrumented by the random treatment assignment. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A22: Overall Effect of the Mentoring Program: Leave-One-Out Estimation

	Aachen	Berlin	Cologne	Duisburg-Essen	Hamburg	Leipzig	Luebeck	Lueneburg	Mannheim
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment	0.451***	0.684***	0.744***	0.524***	0.494***	0.556***	0.513***	0.528***	0.539***
	(0.151)	*(0.160)*	*(0.170)*	*(0.144)*	*(0.154)*	*(0.143)*	*(0.155)*	*(0.140)*	*(0.149)*
Treatment x Higher-SES	-0.694***	-0.909***	-0.943***	-0.719***	-0.713***	-0.750***	-0.480*	-0.782***	-0.743***
	(0.229)	*(0.238)*	*(0.252)*	*(0.223)*	*(0.236)*	*(0.221)*	*(0.266)*	*(0.223)*	*(0.225)*
Higher-SES	0.110	0.197	0.359	0.159	0.231	0.182	0.075	0.173	0.153
	(0.211)	*(0.211)*	*(0.217)*	*(0.206)*	*(0.217)*	*(0.199)*	*(0.208)*	*(0.200)*	*(0.197)*
Outcome in t_0	0.478***	0.432***	0.449***	0.468***	0.415***	0.459***	0.464***	0.489***	0.462***
	(0.085)	*(0.086)*	*(0.093)*	*(0.081)*	*(0.082)*	*(0.078)*	*(0.085)*	*(0.079)*	*(0.078)*
Randomization-pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	249	253	247	284	263	300	252	288	296
R^2	0.767	0.755	0.754	0.755	0.761	0.752	0.754	0.765	0.743
Treatment effect for Higher-SES	-0.243*	-0.224	-0.199	-0.195	-0.219	-0.194	0.033	-0.254*	-0.204
	(0.145)	*(0.148)*	*(0.148)*	*(0.141)*	*(0.146)*	*(0.138)*	*(0.171)*	*(0.143)*	*(0.138)*

Notes: Table shows ITT effects of the mentoring program on the index of labor-market prospects. Site in the column header is dropped from the sample. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in t_0 are included. Robust standard errors in parentheses. Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A23: Effect of the Mentoring Program: Pair in Same Class

	Baseline	Pair in same class	
		Yes	No
	(1)	(2)	(3)
Treatment	0.556***	0.525***	0.789**
	(0.143)	*(0.163)*	*(0.361)*
Treatment x Higher-SES	-0.748***	-0.720***	-0.854
	(0.220)	*(0.246)*	*(0.577)*
Higher-SES	0.182	0.001	0.319
	(0.195)	*(0.257)*	*(0.376)*
Outcome in t_0	0.459***	0.440***	0.474***
	(0.078)	*(0.108)*	*(0.145)*
Randomization-pair fixed effects	Yes	Yes	Yes
Covariates	Yes	Yes	Yes
Observations	304	229	75
R^2	0.753	0.751	0.854
Treatment effect for Higher-SES	-0.192	-0.195	-0.065
	(0.137)	*(0.158)*	*(0.330)*

Notes: Table shows ITT effects of the mentoring program on the index of labor-market prospects. In column 2, sample is restricted to pairs in the same class. In column 3, sample is restricted to pairs not in the same class. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in t_0 are included. Robust standard errors in parentheses. Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A24: Effect of the Mentoring Program on Math Achievement: Pair in Same Class

	Baseline	Pair in same class	
		Yes	No
	(1)	(2)	(3)
Treatment	0.294**	0.379**	-0.061
	(0.142)	*(0.163)*	*(0.345)*
Treatment x Higher-SES	-0.467**	-0.433*	-0.434
	(0.230)	*(0.259)*	*(0.518)*
Higher-SES	0.283	0.179	0.166
	(0.196)	*(0.228)*	*(0.408)*
Outcome in t_0	0.488***	0.579***	0.496***
	(0.100)	*(0.099)*	*(0.159)*
Randomization-pair fixed effects	Yes	Yes	Yes
Covariates	Yes	Yes	Yes
Observations	294	221	73
R^2	0.775	0.762	0.891
Treatment effect for Higher-SES	-0.172	-0.054	-0.495
	(0.145)	*(0.174)*	*(0.351)*

Notes: Table shows ITT effects of the mentoring program on math achievement in school. In Column (2), sample is restricted to pairs in the same class. In Column (3), sample is restricted to pairs not in the same class. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 dimensions. Dummies for missing values in t_0 are included. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A25: Evidence on the Mentoring Relationships

	All	Low-SES	Higher-SES	Difference	
				(2)-(3)	*p*-value
	(1)	(2)	(3)	(4)	(5)
A. Qualitative factors of relationship					
Mentee better at school because of mentor	0.20	0.28	0.14	0.14	0.035
Mentor helped solve non-school-related problems	0.30	0.38	0.23	0.16	0.044
Mentor is role model	0.27	0.32	0.22	0.11	0.153
Parents like that their child has mentor	0.54	0.46	0.61	-0.16	0.060
Mentee had a say in which mentor he/she got	0.47	0.43	0.50	-0.07	0.418
Friends support mentee having a mentor	0.26	0.25	0.27	-0.02	0.822
Mentee and mentor are good friends	0.49	0.51	0.47	0.04	0.622
Mentee satisfied with mentoring relationship	0.58	0.56	0.61	-0.05	0.555
B. Initiation and continuation of relationship					
Mentee has met mentor at least once	0.86	0.82	0.90	-0.08	0.150
Mentoring relationship still exists	0.63	0.56	0.70	-0.15	0.059
Mentoring relationship still exists (conditional on mentor/mentee ever met)	0.73	0.68	0.77	-0.09	0.261
C. Meeting frequency and duration					
Meet at least once per month (in person)	0.50	0.50	0.50	0.00	1.00
Meet at least once per month (all channels)	0.61	0.57	0.66	-0.09	0.256
Duration of meetings (hours)	3.13	2.93	3.31	-0.37	0.386
D. Topics discussed during meetings					
School	0.66	0.64	0.67	-0.03	0.676
Leisure activities	0.57	0.46	0.67	-0.21	0.012
Future in general	0.57	0.57	0.57	0.00	0.995
Occupational and educational future	0.50	0.49	0.51	-0.02	0.808
Personal issues	0.49	0.48	0.50	-0.02	0.795
Family issues	0.25	0.26	0.24	0.02	0.741
Other topics	0.13	0.10	0.16	-0.06	0.313
Don’t know	0.20	0.23	0.17	0.06	0.367
Mentee can decide what is done in meetings	0.62	0.59	0.64	-0.05	0.529

Notes: Table shows group means of variables characterizing the mentoring relationships, based on the follow-up questionnaires of adolescents in the treatment group. Sample: column 1: all respondents (n=153); column 2: low-SES respondents (n=72); column 3: higher-SES respondents (n=81). Column 5 shows the p-value of the coefficient on the higher-SES indicator in a regression of the specific variable on a higher-SES indicator.

Appendix B: Surveying Frame: Sites, Cohorts, and Timing

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

B.1 Selection of Participating Sites

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

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

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

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

B.2 Two-Cohort Sampling Frame

In our pre-analysis plan (contained in the grant application registered with the funding foundations on May 12, 2015), we envisaged a two-cohort sampling procedure to provide a sufficiently large sample to reliably estimate treatment effects. Figure 1 depicts the timeline of the baseline and follow-up surveys in the two cohorts. Appendix Tables A1 and A2 show the dates and sample sizes by cohort and mentoring site for the baseline and follow-up surveys, respectively.

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

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

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

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

Appendix C: Pair-wise Randomization Design

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

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

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

$$\Delta(\boldsymbol{X_i}, \boldsymbol{X_j}) = \sqrt{(\boldsymbol{X_i} - \boldsymbol{X_j})' \boldsymbol{\Sigma^{-1}} (\boldsymbol{X_i} - \boldsymbol{X_j})} \tag{C1}$$

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

As the quality of the balancing for each variable deteriorates as more variables are included in the randomization process, we restricted the set of baseline variables considered in the matching to variables that are expected to both be highly predictive of future outcomes and have

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Appendix D: Attrition Analysis

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

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

In general, attrition is extremely low. In every subsample, we were able to achieve recontact rates of 89 percent or higher. In fact, in almost all cases, recontact rates were above 95 percent. We only observe a slightly smaller recontact rate for low-SES control-group individuals in the survey sample (89 percent).

Results in Table D2 show that attrition in the survey sample is not selective based on observables in the baseline period. The table regresses an attrition indicator on the treatment indicator, the *Higher-SES* indicator, the index of labor-market prospects at the baseline period, and their interactions. Moreover, the table shows that attrition is an issue neither in the overall nor in the administrative samples since all coefficients are close to zero and statistically insignificant.

One feature of the pair-wise randomization design is that we can exclude pairs where (at least) one individual could not be reached in the follow-up, which preserves internal validity (Bruhn and McKenzie (2009)). Table D3 shows the results of this exercise for the index of labor-market prospects. As expected (given the low attrition and the inclusion of randomization-pair fixed effects in the baseline model), there are basically no differences between the results from the main specification in column 1 and the other models, which drop pairs including at least one drop-out in the overall sample (column 2), in the survey sample (column 3), or in the administrative sample (column 4).

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

Table D1: Sample Observations

	Treatment (1)	Control (2)	Total (3)
A. Total sample			
Baseline survey	153	155	308
Follow-up total	152 *(99.4%)*	152 *(98.0%)*	304 *(98.7%)*
Follow-up survey	146 *(95.4%)*	145 *(93.5%)*	291 *(94.5%)*
Follow-up administrative data	146 *(95.4%)*	148 *(95.5%)*	294 *(95.5%)*
B. Low-SES			
Baseline survey	72	73	145
Follow-up total	71 *(98.6%)*	71 *(97.3%)*	142 *(97.9%)*
Follow-up survey	70 *(97.2%)*	65 *(89.0%)*	135 *(93.1%)*
Follow-up administrative data	67 *(93.1%)*	67 *(91.8%)*	134 *(92.4%)*
C. Higher-SES			
Baseline survey	81	82	163
Follow-up total	81 *(100%)*	81 *(98.8%)*	162 *(99.4%)*
Follow-up survey	76 *(93.8%)*	80 *(97.6%)*	156 *(95.7%)*
Follow-up administrative data	79 *(97.5%)*	81 *(98.8%)*	160 *(98.2%)*

Notes: Table shows observation numbers and relative resurvey probabilities (in parentheses) by treatment status, SES background, and sample. Respondents are classified as low-SES if they have at most 25 books at home in the baseline survey and as higher-SES if they have more than 25 books at home at baseline.

Table D2: Attrition Analysis

	Overall sample				Survey sample				Administrative sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treatment	-0.013	-0.013	-0.025	-0.024	-0.022	-0.020	-0.091**	-0.085*	0.000	0.001	0.001	0.006
	(0.009)	*(0.009)*	*(0.019)*	*(0.019)*	*(0.023)*	*(0.024)*	*(0.045)*	*(0.046)*	*(0.018)*	*(0.019)*	*(0.041)*	*(0.045)*
Treatment x Outcome index in t_0		0.000		0.001		0.019		0.023		0.009		0.011
		(0.013)		*(0.023)*		*(0.022)*		*(0.034)*		*(0.023)*		*(0.042)*
Outcome index in t_0		-0.000		0.001		0.006		0.009		-0.001		0.025
		(0.012)		*(0.025)*		*(0.023)*		*(0.034)*		*(0.024)*		*(0.039)*
Treatment x Higher-SES			0.022	0.021			0.131**	0.122*			-0.000	-0.012
			(0.021)	*(0.021)*			*(0.065)*	*(0.066)*			*(0.051)*	*(0.057)*
Higher-SES			-0.027	-0.027			-0.070	-0.062			-0.033	-0.025
			(0.020)	*(0.020)*			*(0.043)*	*(0.046)*			*(0.034)*	*(0.038)*
Treatment x Higher-SES x Outcome index in t_0				-0.008				-0.022				-0.029
				(0.020)				*(0.047)*				*(0.046)*
Higher-SES x outcome index in t_0				-0.001				-0.007				-0.039
				(0.024)				*(0.043)*				*(0.041)*
Randomization-pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	No	No	No	No	No	No	No	No	No	No	No
Observations	308	308	308	308	308	308	308	308	308	308	308	308
R^2	0.75	0.75	0.75	0.75	0.59	0.59	0.60	0.60	0.70	0.70	0.70	0.71
Attrition *F*-test *p*-values												
Overall sample	0.151	0.361	0.355	0.730	0.347	0.478	0.117	0.392	1.000	0.922	1.000	0.765
Low-SES sample			0.190	0.304			0.045	0.291			0.989	0.792
Higher-SES sample			0.708	0.485			0.241	0.413			0.998	0.285

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

Table D3: Effect of the Mentoring Program: Dropping Attrition Pairs

	Baseline	Dropping pairs in		
		Overall sample	Survey sample	Administrative sample
	(1)	(2)	(3)	(4)
Treatment	0.556***	0.556***	0.583***	0.556***
	(0.143)	*(0.142)*	*(0.151)*	*(0.153)*
Treatment x Higher-SES	-0.748***	-0.748***	-0.779***	-0.784***
	(0.220)	*(0.219)*	*(0.232)*	*(0.234)*
Higher-SES	0.182	0.182	0.167	0.155
	(0.195)	*(0.194)*	*(0.206)*	*(0.197)*
Outcome in t_0	0.459***	0.459***	0.451***	0.470***
	(0.078)	*(0.077)*	*(0.082)*	*(0.082)*
Randomization-pair fixed effects	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes
Observations	304	302	277	286
R^2	0.465	0.468	0.451	0.470
Treatment effect for Higher-SES	-0.192	-0.192	-0.196	-0.227
	(0.137)	*(0.136)*	*(0.143)*	*(0.141)*

Notes: Table shows ITT effects of the mentoring program on the index of labor-market prospects. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in t_0 are included. Robust standard errors in parentheses. Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1.

Appendix E: Treatment Effects by Alternative SES Dimensions

This appendix reports results on heterogeneous treatment effects by migrant status (Appendix E.1), by single-parenthood status (Appendix E.2), and by a broader SES measure (Appendix E.3).

E.1 Heterogeneous Treatment Effects by Migrant Status

Ample evidence indicates that migrants usually fare worse in terms of economic outcomes than the native population (e.g., Algan et al. (2010)). Also in our sample, we observe that the migrant share is higher among low-SES adolescents (72 percent) than among higher-SES adolescents (45 percent). However, having a migrant background does not necessarily imply low-SES background (e.g., when considering books at home or parental education and employment), which raises the question whether the mentoring program is more effective for migrants than for natives on average.

Table E1 shows a specification that interacts treatment with migrant status. The non-migrant indicator takes a value of one if the respondent and both parents were born in Germany, zero otherwise. Column 2 shows that the mentoring program has a significant positive effect on the index of labor-market prospects for migrants on average. However, the effect is only half as large as for low-SES adolescents (29.0 vs. 55.6 percent of a standard deviation, see column 1). Still, with a native-migrant gap in labor-market prospects of 22.4 percent of a standard deviation (see bottom of Table E1), the mentoring program is able to fully close the gap. The mentoring program does not significantly affect the index of labor-market prospects for natives on average. Column 3 includes the treatment interactions with both higher-SES and non-migrant status to examine whether the migrant heterogeneity has explanatory power over and above the SES heterogeneity. The treatment interaction with the non-migrant status becomes very small and statistically insignificant in this model, supporting our choice of baseline SES measure.[8]

Examining the components of the index of labor-market prospects in the odd-numbered columns of Table E2, we consistently find positive treatment effects for migrants, but they are significant only for labor-market orientation. This is particularly interesting given that there is a sizeable native-migrant gap in labor-market orientation of 16 percent of a standard deviation.

[8] Throughout, we generally refrain from estimating models with triple interactions between treatment, SES, and a third dimension, which tend to become very imprecise.

There is no such gap between low-SES and higher-SES individuals (see Table 5). A possible explanation is that migrants are less familiar with the German apprenticeship system, which does not exist or is very differently organized in most other countries (e.g., Kristen, Reimer, and Kogan (2008)). The mentoring program is able to close the gap in labor-market orientation entirely.

The migrant status includes both first-generation migrants (respondent born abroad; 22% of all migrants) and second-generation migrants (respondent born in Germany, but at least one of the respondent's parents born abroad; 78% of all migrants). In the even-numbered columns of Table E2, we study treatment effect heterogeneity by more detailed migrant status. Column 2 reveals that the mentoring program has a very large effect of 73 percent of a standard deviation for first-generation migrants and only a very modest effect of 13.4 percent of a standard deviation for second-generation migrants. Thus, the program is considerably more effective for first-generation migrants than for second-generation migrants or natives. We also find substantially larger treatment effects for first- compared to second-generation migrants on math achievement (column 4) and patience and social skills (column 6), while the treatment effect on labor-market orientation is rather similar for first- and second-generation migrants (column 8).

E.2 Heterogeneous Treatment Effects by Single-Parenthood Status

In the baseline survey, we asked the adolescents with whom they usually live together. When not distinguishing between biological parents and step-parents, 25 percent of adolescents in our sample (30 percent in the low-SES sample and 21 percent in the higher-SES) report that at least one parent is absent. If we consider only biological parents, this share is even larger at 36 percent (41 percent in the low-SES sample and 30 percent in the higher-SES sample). According to official statistics from the Federal Statistical Office for the year 2017, 19 percent of families with children of all ages have a single-parenthood status. In PISA 2012, 14 percent of adolescents have a single parent (17 percent for low-SES and 13 percent for higher-SES). Adolescents in our sample are thus much more likely to live in single-parent households than the average adolescent in Germany.

Table E3 examines effect heterogeneity by single-parenthood status. The non-single-biological-parent indicator takes a value of one if the adolescent lives together with both biological parents and zero otherwise. The non-single-parent indicator takes a value of one if the adolescent lives together with two parents (either biological parents or step parents) and zero

otherwise. The first two columns show that the average treatment effect for pupils with a single parent is positive, but not statistically significant at conventional levels.[9] When jointly including the interactions of treatment with higher-SES and non-single parenthood in columns 3 and 4, the low-SES status is much more relevant for the treatment effect to materialize than having a single parent per se.

E.3 Heterogeneous Treatment Effects for a Broader SES Measure

In our main specification, we measure SES by an indicator of books at home. In the background survey, we also collected information on other potential proxies for SES status, namely parental education (a dummy variable taking a value of one if the parent obtained a university degree and zero otherwise) and parental employment (distinguishing between full-time employed, part-time employed, not employed, and unemployed).[10] Appendix Table A4 shows the distributions of these alternative SES proxies in our sample. The table also reveals a high number of missing values in these variables: the share of missing values is 40 (32) percent for fathers' (mothers') education and 19 (9) percent for fathers' (mothers') employment. These missing shares exceed the respective shares in the average population to a considerable extent. For instance, compared to shares of missing values in PISA 2012, the shares in our sample are 26 (22) percentage points higher for fathers' (mothers') education and 12 (5) percentage points higher for fathers' (mothers') employment. This likely reflects that the target population of the mentoring program are low-SES individuals who are often unaware of the educational and employment background of their parents. These high shares of missing values advocate the use of books at home – which does not have any missing values in our sample – as the SES measure in our main analysis.

Despite the shortcomings of the additional SES proxies, we use the information on parental education and employment to construct a broader measure of SES background that does not rely on books at home alone.

[9] The treatment effect for adolescents with a single biological parent in column 1 is significant at the 13 percent level (p=0.127).

[10] Note that the ordering of the employment status indicators is in accordance with the expected direction of prediction for a higher SES status. In that sense, we expect that being unemployed (and searching for a job) corresponds to the lowest SES status. Examining the percentile positions in the ESCS distribution in PISA 2012 confirms this expectation: Individuals with full-time employed, part-time employed, not employed (not searching), and unemployed fathers come from the 52nd, 45th, 43rd, and 31th percentiles, respectively.

In a first step, we impute the missing values in the parental university dummies and the employment categories. We do so by using data from PISA 2012. Separately for fathers and mothers, we estimate the relationship between education/employment and the PISA index of economic, social, and cultural status (ESCS).[11] As both parental-background variables enter the construction of the index, they are strongly associated with ESCS (Table E4). Making use of this relationship, we predict parental education and employment status for those individuals who did not report this information in PISA. We use these average predictions to characterize the population with missing values in our dataset. Applying this imputation procedure, we predict that 4.9 (3.0) percent of fathers (mothers) of individuals in our sample who did not report parental education have a university degree (compared to 23 (18) percent on average). For the employment status, the predicted average for missing values in the linear employment index (ranging from 0=unemployed to 3=full-time employed)[12] is 2.71 (2.05) for fathers' (mothers') employment (compared to 2.77 (2.06) on average).[13]

In the next step, we construct a PISA-based SES index using principal component analysis (PCA) on the PISA data to combine books at home, fathers' and mothers' university degree, and fathers' and mothers' employment status. The SES index is the first principal component of these variables. Table E5 shows that this index is highly correlated with books at home, but also with the ESCS as a broader measure of socioeconomic background. The table also shows that the correlation with the math test score in PISA is very similar across the different SES indices.

In the final step, we construct the same SES index in our dataset. To avoid the use of potentially endogenous weights, we use the factor loadings from the PCA on the PISA data for

[11] As a composite measure of students' socioeconomic background in PISA, the ESCS index is based on a principal component analysis of three inputs: highest occupational status of parents, highest education level of parents, and home possessions (which includes books at home; see OECD (2013), p. 263).

[12] The linear employment variable takes the following values: 0=unemployed, searching for a job; 1=not employed, not searching for a job; 2=part-time employed; 3=full-time employed.

[13] The prediction results on parental education suggest that individuals with missing information in this category have a particularly low SES background. This conjecture is corroborated when examining the math achievement for individuals with missing information on fathers' education (results for missing mother's education are very similar). They score on average 468 PISA points (rank at the 35th percentile) in math performance (compared to 528 PISA points (rank at the 53rd percentile) for those without missing values), which represents a substantial difference in achievement (of about 1.5 school years). We can also check how often individuals with missing information on fathers' education belong to the lowest two books-at-home categories, which corresponds to our preferred low-SES definition of having at most 25 books at home. Interestingly, in the PISA data 40 percent of individuals with missing father education belong to this group – a 20 percentage-point gap compared to the average population.

the construction of the extended SES index in our data. This also requires the standardization of variables with averages and standard deviations from the PISA sample. Defining individuals as low-SES when they are in the lowest tercile in the distribution of the PISA-based SES index yields a low-SES share of 44 percent in our sample, which is similar to the low-SES share based on books at home. That is, the PISA-based SES index shows once again that our sample is predominately drawn from the lower part of the SES distribution. Table E6 shows the correlation between the PISA-based SES index and books at home in our sample. As expected, the correlation is high, but far from perfect as more than one-quarter of our sample (26 percent) is categorized into a different SES group.

Using the extended SES background index to define low-SES and higher-SES adolescents, Table E7 reports the results on treatment effects of the mentoring program on the index of labor-market prospects (column 1) and its three components (columns 2-4). The results based on the broader SES measure confirm that the mentoring program has significant effects only for low-SES adolescents.

When comparing the results for low-SES respondents defined using the broader SES index to the low-SES results from our main specification based on books at home, the overall results look very similar. Treatment effects on the index of labor-market prospects (column 1 of Table E7 and column 4 of Table 2) are almost identical for both SES definitions. Interestingly, the effect on math grades is slightly stronger for the broader SES index (column 2 of Table E7 vs. column 1 of Table 3), which may be due to the fact that it includes parental education. In turn, the treatment effect on the patience and social skills index is smaller when using the broader SES index (column 3 of Table E7 and column 1 of Table 4). However, the treatment effect on patience is almost unchanged when compared to the results from the main specification (not shown). Treatment effects on labor-market orientation are again similar for both SES definitions. However, the treatment effect in column 4 of Table E7 is less precisely estimated than the corresponding estimate in column 1 of Table 5 and is just shy of statistical significance at conventional levels (*p*-value: 0.11). In general, standard errors are slightly larger with the more encompassing SES measure for each of the outcomes, which is likely due to an increase in the measurement error resulting from predicting missing values on parental education and employment from the PISA data when deriving the SES index.

Table E1: Effect of the Mentoring Program by Migrant Status

	(1)	(2)	(3)
Treatment	0.556***	0.290**	0.549***
	(0.143)	*(0.122)*	*(0.146)*
Treatment x Higher-SES	-0.748***		-0.765***
	(0.220)		*(0.259)*
Treatment x Non-migrant		-0.328	0.039
		(0.230)	*(0.271)*
Higher-SES	0.182		0.188
	(0.195)		*(0.194)*
Non-migrant	0.177	0.323*	0.159
	(0.160)	*(0.193)*	*(0.201)*
Outcome in t_0	0.459***	0.449***	0.459***
	(0.078)	*(0.083)*	*(0.078)*
Randomization-pair fixed effects	Yes	Yes	Yes
Covariates	Yes	Yes	Yes
Observations	304	304	304
R^2	0.753	0.734	0.753
Treatment effect for Higher-SES	-0.192		-0.216
	(0.137)		*(0.220)*
Treatment effect for Non-migrant		-0.038	
		(0.170)	
Treatment effect for Non-migrant, Low-SES			0.588**
			(0.276)
Migrant gap		0.224	

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

Table E2: Effect of the Mentoring Program by Migrant Status: First- and Second-Generation Migrants

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

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

Table E3: Effect of the Mentoring Program by Single-Parenthood Status

	(1)	(2)	(3)	(4)
Treatment	0.320	0.240	0.667***	0.562**
	(0.209)	*(0.228)*	*(0.219)*	*(0.238)*
Treatment x Higher-SES			-0.729***	-0.776***
			(0.224)	*(0.224)*
Treatment x Non-single biological parent	-0.253		-0.185	
	(0.292)		*(0.274)*	
Treatment x Non-single parent		-0.127		-0.005
		(0.284)		*(0.267)*
Higher-SES			0.163	0.202
			(0.200)	*(0.198)*
Both biological parents	0.165		0.099	
	(0.202)		*(0.196)*	
Both parents		-0.034		-0.158
		(0.228)		*(0.212)*
Outcome in t_0	0.450***	0.442***	0.465***	0.460***
	(0.082)	*(0.083)*	*(0.078)*	*(0.079)*
Randomization-pair fixed effects	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes
Observations	304	304	304	304
R^2	0.732	0.732	0.754	0.755
Treatment effect for Higher-SES			-0.063	-0.214
			(0.230)	*(0.238)*
Treatment effect for Non-single parent	0.067	0.113		
	(0.137)	*(0.116)*		
Treatment effect for Non-single parent, Low-SES			0.482***	0.557***
			(0.177)	*(0.165)*

Notes: Table shows ITT effects of the mentoring program on the index of labor-market prospects. *Non-single biological parent* is a dummy variable that is one if the adolescent lives together with both biological parents and zero otherwise. *Non-single parent* is a dummy variable that is one if the adolescent lives together with both a biological or step-father and a biological or step-mother, and zero otherwise. Covariates are from the baseline survey and include: age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table E4: Association of ESCS with Parental Education and Employment in PISA

	University degree		Employment status	
	Father	Mother	Father	Mother
	(1)	(2)	(3)	(4)
ESCS	0.312***	0.222***	0.100***	0.097***
	(0.006)	*(0.007)*	*(0.012)*	*(0.014)*
R^2	0.406	0.264	0.021	0.012
Observations	3,586	3,734	3,872	3,969

Notes: Table shows correlations of the index of economic, social, and cultural status (ESCS) with the dummy of holding a university degree (columns 1 and 2) and with the linear index of the employment status (categories: 0=unemployed, searching for a job; 1=not employed, not searching for a job; 2=part-time employed; 3=full-time employed) (columns 3 and 4). Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Data source: PISA 2012.

Table E5: Correlation of PISA-based SES Index with Other SES Measures in PISA

	PISA-based SES index	Books at home	ESCS	Math score
	(1)	(2)	(3)	(4)
PISA-based SES index	1.000			
Books at home	0.661	1.000		
ESCS	0.750	0.531	1.000	
Math score	0.386	0.434	0.414	1.000

Notes: Table shows correlations of the PISA-based SES index with other indices of SES background and math test score. N=4,078. Data source: PISA 2012.

Table E6: Correlation of PISA-based SES Index and Books at Home in Our Sample

	SES indices		Low-SES indices	
	PISA-based SES index	Books at home	PISA-based SES index	Books at home
	(1)	(2)	(3)	(4)
SES indices				
PISA-based SES index	1.000			
Books at home	0.652	1.000		
Low-SES indices				
PISA-based SES index	-0.895	-0.558	1.000	
Books at home	-0.507	-0.818	0.484	1.000

Notes: Table shows correlations of the PISA-based SES index with books at home in the baseline sample (*N*=308). *SES indices* are the linear indices of the PISA-based SES index and the books-at-home index. Higher values in these indices refer to higher SES background. *Low-SES indices* indicate the low-SES population of each index. PISA-based SES index: lowest tercile in the PISA distribution. Books at home: at most 25 books at home.

Table E7: Effect of the Mentoring Program: Broader SES Measure

	Outcome index	Math grade	Patience and social skills index	Labor-market orientation index
	(1)	(2)	(3)	(4)
Treatment	0.569***	0.478***	0.286*	0.303
	(0.152)	*(0.146)*	*(0.170)*	*(0.188)*
Treatment x Higher-SES	-0.745***	-0.787***	-0.289	-0.305
	(0.236)	*(0.221)*	*(0.277)*	*(0.302)*
Higher-SES	0.330*	0.497***	0.115	-0.025
	(0.186)	*(0.177)*	*(0.225)*	*(0.210)*
Outcome in t_0	0.445***	0.536***	0.259**	0.358***
	(0.080)	*(0.097)*	*(0.105)*	*(0.091)*
Randomization-pair fixed effects	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes
Observations	304	294	291	291
R^2	0.748	0.789	0.685	0.694
Treatment effect for Higher-SES	-0.175	-0.308**	-0.003	-0.002
	(0.140)	*(0.132)*	*(0.176)*	*(0.174)*
SES gap	0.289	0.394	0.214	-0.084

Notes: Table shows ITT effects of the mentoring program on the outcome indicated in the column header. *Higher-SES* refers to the extended PISA-based SES index, which comprises books at home, university education of the father/mother, and employment status of the father/mother. *SES gap* is calculated on the control-group sample in the follow-up survey as the coefficient on higher-SES background in a regression of the respective outcome on the higher-SES indicator. Covariates are from the baseline survey and include: gender, migrant, age, received paid private teaching, parental homework support, and Big-5 personality traits. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

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

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

F.1 The PIAAC and PIAAC-L Data

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

PIAAC was designed to measure key cognitive and workplace skills needed for individuals to advance in their jobs and participate in society. The survey assessed cognitive skills in three domains: numeracy, literacy, and problem-solving in technology-rich environments. The domains refer to key information-processing competencies. Numeracy skills are defined as the ability to access, use, interpret, and communicate mathematical information and ideas in order to engage in and manage the mathematical demands of a range of situations in adult life. Literacy skills are defined as the ability to understand, evaluate, use and engage with written texts to participate in society, to achieve one's goals, and to develop one's knowledge and potential. The domain of problem-solving in technology-rich environments, typically referred to as "ICT skills," is defined as the ability to use digital technology, communication tools, and networks to acquire and evaluate information, communicate with others, and perform practical tasks.[14] In the empirical analysis, test scores in each domain are standardized with a mean of zero and a standard deviation of one.

[14] Not all respondents participated in the ICT-skills assessment, because of a lack of any computer experience, failing a short initial ICT test, or opting out of the domain (see Falck, Heimisch-Roecker, and Wiederhold (2020) for details).

Germany conducted a follow-up study, PIAAC-L, in which respondents who participated in the original German PIAAC study in 2011/2012 were interviewed in three further waves (2014, 2015, and 2016).[15] For this analysis, we focus on the first wave of PIAAC-L, which elicited more detailed information from the participants regarding their educational history, personality traits, and family background. In particular, respondents reported the grades in mathematics, German, and the first foreign language (typically English) from their last report card in secondary school.

F.2 School Grades and Later-Life Outcomes

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

In Table F1, we show how school grades are related to important adult outcomes. Columns 1-6 consider cognitive skills in numeracy, literacy, and ICT. Columns 7-12 focus on labor-market outcomes, investigating unemployment (columns 7-8) as well as monthly and hourly wages (columns 9-12).[16] Following our main specification to evaluate the impact of the mentoring program, we interact grades with an indicator for higher-SES (1: more than 25 books at home at age of 16; 0: otherwise) to investigate whether grade effects differ by SES background. Regressions control for demographic characteristics (a quadratic polynomial in age, gender, and migration status)[17] as well as school-type fixed effects. The grade scale is reversed, such that better grades indicate more beneficial outcomes, and standardized with a mean of zero and a standard deviation of one (normalized to the distribution of the estimation sample in column 1 of Table F1). The odd columns in the table include only math grades, the primary cognitive outcome measure in our evaluation study, in addition to the controls. The even columns also include German and foreign-language grades.[18]

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

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

[17] Migration status indicates whether a respondent was born in Germany (i.e., first-generation migrant).

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

The results in Table F1 show a clear pattern: math grades at the end of secondary school are a significant predictor of cognitive skills and labor-market success later in life. As expected, math grades are more strongly correlated with numeracy skills than with literacy and ICT skills in adulthood, but estimates are sizeable for all three cognitive outcomes. For respondents with a low-SES background, an improvement in math grades by one standard deviation is related to an increase in adult numeracy skills by 17 percent of a standard deviation (column 1), in adult literacy skills by 13 percent of a standard deviation (column 3), and in adult ICT skills by 12 percent of a standard deviation (column 5). The relationship between math grades and labor-market success is also strong. When math grades increase by one standard deviation, the unemployment rate of respondents with a low-SES background decreases by 1.2 percentage points (26 percent of the full-sample mean and 22 percent of the mean in the low-SES sample) (column 7), while their monthly wages increase by 7.8 percent (column 9) and their hourly wages by 4.1 percent (column 11).[19] Grade effects do not differ significantly by SES background, as the interaction between grades and the higher-SES indicator is small and typically insignificant.

When we also include German and foreign-language grades at school (even columns of Table F1), the math-grade estimates are barely affected. Most strikingly, German grades and foreign-language grades are only weakly, if at all, related to cognitive skills and labor-market success in adulthood when math grades are also included. While the coefficients on German grades are small and insignificant across all outcomes, foreign-language grades are modestly related to cognitive skills, but play no role for labor-market outcomes. These results indicate that math grades at school are far more relevant in predicting human-capital formation and labor-market success later in life than German or foreign-language grades. This provides a strong argument for focusing on math grades as a proxy for cognitive skills in the experimental analysis of the mentoring program.

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

F.3 Behavioral Traits and Labor-Market Outcomes

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

Unfortunately, PIAAC-L did not assess individuals' patience directly. However, the concept of grit is strongly related to patience, as it is defined as "perseverance and passion for long-term goals" (Duckworth et al. (2007)). In Table F2, we link the PIAAC-L measure of grit[20] to labor-market outcomes assessed in PIAAC.[21] As in Table F1, we consider unemployment as well as monthly and hourly wages. In the odd columns, grit (as well as its interaction with an indicator for higher-SES background) is included together with standard demographic controls. In the even columns, we add other personality traits as further controls.[22]

Across specifications, grit is strongly related to labor-market outcomes. For respondents with a low-SES background, a one-standard-deviation increase in grit is related to a decrease in unemployment by 2.6 percentage points (57 percent of the full-sample mean and 46 percent of the low-SES-sample mean) (column 1), an increase in monthly wages by 9 percent (column 3), and an increase in hourly wages by 5.5 percent (column 5). Grit effects on unemployment and hourly wages tend to be somewhat stronger for individuals with low-SES background, albeit not significantly so. When adding the other personality traits in the even columns, the grit coefficients even tend to increase. Among the other personality traits, higher values of trust (see below) and external locus of control are consistently related to better labor-market outcomes, and

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

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

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

so are lower values of extraversion and agreeableness. However, with the exception of trust, effect sizes are smaller than those for grit.

There is little prior evidence on the importance of trust for individuals' labor-market outcomes. One noticeable exception is the work by Butler, Giuliano, and Guiso (2016), who find a non-linear relationship between trust and household income in the European Social Survey (ESS). In the ESS, trust is measured using the question: "Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?" with answer categories on a scale from zero to ten. The authors find that for trust levels between zero and seven, an increase in trust is associated with higher household income; for higher levels of trust, more trust is associated with a decrease in income. In our sample of applicants to the mentoring program, the average baseline level of trust is below the "critical" value identified by Butler, Giuliano, and Guiso (2016), at 6.2 in the treatment group and 6.3 in the control group.[23] Aggregate evidence supports a positive relationship between trust and income at the country level (Knack and Keefer (1997); Algan and Cahuc (2010)).

In Table F3, we assess the relationship between trust and individual labor-market outcomes.[24] The table is constructed analogously to Table F2. Across specifications, trust is strongly related to labor-market outcomes. For respondents with a low-SES background, a one-standard-deviation increase in trust is related to a decrease in unemployment by 2.2 percentage points (48 percent of the full-sample mean and 39 percent of the low-SES-sample mean) (column 1), an increase in monthly wages by 16.2 percent (column 3), and an increase in hourly wages by 8.1 percent (column 5). The interaction of trust with the higher-SES indicator suggests that trust effects do not differ significantly by SES background. When the other personality traits are included in the even columns, the trust coefficient decreases somewhat, but remains statistically significant. In contrast to Butler, Giuliano, and Guiso (2016), we do not find evidence for a hump-shaped relationship between trust and wages (not shown).

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

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

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

F.4 Professional Qualifications by SES Background and School Grades

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

We start with evidence supporting the idea that successfully completing an apprenticeship is a desirable outcome for the target group of the mentoring program. The upper panel of Table F4 documents a substantial SES gap in the probability of failing to obtain any professional qualification, i.e., obtaining neither an apprenticeship nor a university degree. Focusing on those aged over 35 years (who are likely to have completed their final educational degree), 20 percent of individuals with low-SES background have no professional qualification, compared to only 5 percent in the group of higher-SES individuals. Results are very similar when considering individuals aged over 25, 30, or 40 years, suggesting strong persistence over the lifecycle and thus a policy focus on alleviating SES differences early in the professional career.[25]

The large SES gap in successfully obtaining a professional qualification is partially due to the fact that individuals with a low-SES background are more likely to drop out of an apprenticeship than their higher-SES counterparts. Focusing on persons older than 30 years, most of whom have finished their formal education, the probability of individuals with low-SES background to ever have dropped out of apprenticeship training is 7.9 percent, compared to 4.5 percent for individuals with a higher-SES background (middle panel of Table F4).[26] From those individuals having experienced an apprenticeship dropout, almost two-thirds (65 percent) in the low-SES sample have not obtained any professional qualification by the age of 31, compared to one-third in the higher-SES sample. These differences remain considerable even when acknowledging that the share of individuals with low-SES background who have completed an apprenticeship by the age of 31 is somewhat larger than the corresponding share for individuals with a higher-SES background (71 vs. 67 percent).

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

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

At the same time, only 6 percent of individuals with a low-SES background have obtained a university degree by the age of 31 – compared to 26 percent of individuals with higher-SES background (bottom panel of Table F4). Further taking into account that 2 percent of individuals in the low-SES group have experienced a university dropout (of whom 87 percent have not obtained a university degree) by the age of 31,[27] the evidence suggests that university education is not a viable option for the overwhelming majority of individuals with low-SES background.

This evidence has important implications for the qualification outcomes to be considered in the evaluation of the mentoring program. The mentoring program is not designed to address the (apparently substantial) barriers to enter university for disadvantaged youths, which likely include lacking educational aspirations of parents, peer effects being absent or even negative due to low-ability peers, low school quality, and others. Moreover, since the mentoring program is targeted towards adolescents from lower-track secondary schools, even successfully finishing these schools does not provide a university entrance qualification. Accordingly, only 5 percent of individuals who obtained their highest school-leaving certificate from low-track (*Hauptschulen*) or intermediate-track (*Realschulen*) secondary schools successfully completed university education by the age of 31. For individuals with a low-SES background, this share is only 2 percent (see bottom panel of Table F5).[28] Thus, entering university is simply no option for the vast majority of low-SES participants in the mentoring program, at least not in the short run. In the context of our study, the question is rather whether the mentoring program can help disadvantaged youths to find an apprenticeship after school and to successfully complete it.[29]

In the paper, we document a strong effect of the mentoring program on math grades for low-SES participants. The linked PIAAC and PIAAC-L data allow us to investigate whether better grades at the end of secondary school are associated with better qualification outcomes in adulthood. This complements the analysis in Appendix F.2 of grade effects on employment and

[27] Six percent of respondents with higher-SES background have experienced university dropout (of whom 66 percent have not obtained a university degree) by the age of 31.

[28] For respondents who are not currently in the formal education system, PIAAC and PIAAC-L collect information only on the *highest* secondary school degree. Therefore, we cannot observe whether individuals attended a lower-track secondary school before finishing a higher school track. Our sample of individuals with lower-track secondary education as their highest secondary school degree thus likely contains less able graduates from lower-track schools, and thus Table F5 likely underestimates the probability of completing university for the entire population of graduates from lower-track schools.

[29] Mentoring can potentially also help disadvantaged youths to avoid unemployment early in the career. The unemployment rate below the age of 25 is 13.1 percent for individuals in PIAAC with low-SES background and only 7.4 percent for individuals with higher-SES background ($p = 0.084$).

wages. We keep only individuals above the age of 30 to ensure that most of them have finished their formal education. Table F6 is organized analogously to Table F1: While the odd columns include math grades as the only grade variable, the even columns further add German and foreign-language grades. Outcomes are indicators of not having obtained any qualification (columns 1 and 2), of having successfully completed an apprenticeship (columns 3 and 4), and of having quit one or more apprenticeships (columns 5 and 6).

Results in Table F6 show that better math grades decrease the probability of both not having obtained any qualification and having experienced an apprenticeship dropout by the age of 31. In terms of magnitude, a one-standard-deviation increase in math achievement reduces the probability of not having obtained a qualification by 2.5 percentage points for individuals with low-SES background (36 percent of the full-sample mean and 16 percent of the mean in the low-SES sample) (column 1). The probability of apprenticeship dropout is reduced by 1.2 percentage points (24 percent of the full-sample mean and 15 percent of the mean in the low-SES sample) (column 5). Better math grades tend to be negatively related to the probability of completing an apprenticeship, but coefficients are small and at most marginally significant. There is no evidence for SES heterogeneity in grade effects. Furthermore, neither German grades nor foreign-language grades are systematically related to qualification outcomes, conditional on math grades.

As the mentoring program is targeted towards lower-track secondary schools, we also investigate whether better grades improve the career prospects of individuals who graduated from these schools. Table F7 restricts the sample to individuals who obtained their highest school degree from low-track (*Hauptschulen*) or intermediate-track (*Realschulen*) secondary schools. In this sample, math grade effects on qualification outcomes are even stronger than in the full sample. In particular, a one-standard-deviation-increase in math achievement decreases the probability of not having obtained any qualification by the age of 31 for individuals with a low-SES background by 4 percentage points, which corresponds to 57 percent of the mean in the lower-track secondary school sample (36 percent of the low-SES mean) (column 1). Better math grades also increase the probability of finishing an apprenticeship, although the effect magnitude is rather small (column 3). The probability to drop out of an apprenticeship training decreases in math grades, by 1.6 percentage points (27 percent of the sample mean, 20 percent of the low-SES mean) for a one-standard-deviation increase in math achievement (column 5). There is no

apparent effect heterogeneity by SES background, although math grade effects tend to be somewhat stronger for individuals with a higher-SES background once grades in German and foreign language are also included (at least for *No qualification* and *Apprenticeship completion*). Again, these other grades are themselves not significantly related to qualification outcomes in the sample of lower-track school graduates.

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

Table F1: Math Grades at School and Later-Life Outcomes

	Numeracy skills		Literacy skills		ICT skills		Unemployed		Monthly wage		Hourly wage	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Math grade	0.170***	0.168***	0.127***	0.120***	0.124***	0.119***	-0.012***	-0.012***	0.078***	0.076***	0.041***	0.044***
	(0.013)	*(0.014)*	*(0.014)*	*(0.014)*	*(0.015)*	*(0.016)*	*(0.004)*	*(0.004)*	*(0.016)*	*(0.017)*	*(0.010)*	*(0.010)*
Math grade x Higher-SES	0.005*	0.003	0.003	0.003	0.001	0.001	0.001	0.001	0.004	0.001	0.003*	0.002
	(0.003)	*(0.003)*	*(0.002)*	*(0.003)*	*(0.003)*	*(0.004)*	*(0.001)*	*(0.001)*	*(0.003)*	*(0.003)*	*(0.002)*	*(0.002)*
Higher-SES	0.373***	0.349***	0.418***	0.406***	0.302***	0.307***	-0.003	-0.005	0.091**	0.069	0.074**	0.080***
	(0.037)	*(0.039)*	*(0.036)*	*(0.039)*	*(0.042)*	*(0.045)*	*(0.011)*	*(0.011)*	*(0.040)*	*(0.042)*	*(0.024)*	*(0.026)*
German grade		-0.008		0.009		-0.003		-0.004		0.032*		0.007
		(0.015)		*(0.015)*		*(0.017)*		*(0.005)*		*(0.019)*		*(0.012)*
German grade x Higher-SES		0.001		0.000		0.001		0.000		0.003*		0.001
		(0.002)		*(0.002)*		*(0.002)*		*(0.000)*		*(0.002)*		*(0.001)*
Foreign-language grade		0.039**		0.038**		0.034*		0.004		-0.014		-0.016
		(0.016)		*(0.016)*		*(0.018)*		*(0.005)*		*(0.019)*		*(0.012)*
Foreign-language grade x Higher-SES		0.000		-0.000		-0.001		0.000		0.000		-0.001
		(0.001)		*(0.001)*		*(0.001)*		*(0.000)*		*(0.001)*		*(0.001)*
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School-type fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade imputation dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,758	3,758	3,758	3,758	3,228	3,228	3,019	3,019	2,637	2,637	2,419	2,419
R^2 (adjusted)	0.407	0.412	0.407	0.410	0.344	0.345	0.030	0.034	0.266	0.268	0.353	0.356

Notes: Ordinary least squares estimates. Dependent variables: numeracy skills (columns 1 and 2), literacy skills (columns 3 and 4), ICT skills (columns 5 and 6), dummy for unemployment (columns 7 and 8), log monthly wages (columns 9 and 10), and log hourly wages (columns 11 and 12). Cognitive skills and grades are standardized to have mean zero and standard deviation one. *High-SES* indicates whether the respondent had more than 25 books at home at the age of 16. Sample: respondents aged 16-65 years in PIAAC who participated in PIAAC 2011/2012 and PIAAC-L 2014. All specifications control for a quadratic polynomial in age, gender, migration status, as well as for fixed effects for the type of secondary school that respondents attended and imputation dummies for school grades. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Data sources: PIAAC 2011/2012, PIAAC-L 2014.

Table F2: Grit and Labor-Market Outcomes

	Unemployed		Monthly wages		Hourly wages	
	(1)	(2)	(3)	(4)	(5)	(6)
Grit	-0.026**	-0.022*	0.090***	0.117***	0.055***	0.073***
	(0.011)	*(0.011)*	*(0.035)*	*(0.034)*	*(0.020)*	*(0.020)*
Grit x Higher-SES	0.018	0.019	0.036	0.024	-0.016	-0.011
	(0.012)	*(0.012)*	*(0.039)*	*(0.038)*	*(0.024)*	*(0.023)*
Higher-SES	-0.013	-0.005	0.202***	0.137***	0.172***	0.110***
	(0.010)	*(0.011)*	*(0.039)*	*(0.038)*	*(0.024)*	*(0.023)*
Openness		0.001		-0.020		0.016
		(0.005)		*(0.017)*		*(0.011)*
Conscientiousness		-0.008		-0.018		-0.046***
		(0.005)		*(0.020)*		*(0.012)*
Extraversion		0.009**		-0.041**		-0.041***
		(0.004)		*(0.017)*		*(0.010)*
Agreeableness		0.012***		-0.075***		-0.046***
		(0.004)		*(0.016)*		*(0.010)*
Neuroticism		0.013**		0.003		-0.011
		(0.005)		*(0.018)*		*(0.011)*
Trust		-0.013***		0.126***		0.093***
		(0.004)		*(0.017)*		*(0.010)*
Internal locus of control		0.004		-0.014		-0.006
		(0.004)		*(0.017)*		*(0.010)*
External locus of control		-0.010**		0.082***		0.070***
		(0.004)		*(0.018)*		*(0.011)*
Risk attitude		0.007		0.007		-0.027***
		(0.005)		*(0.017)*		*(0.010)*
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Imputation dummies for personality traits	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,019	3,019	2,637	2,637	2,419	2,419
R^2 (adjusted)	0.019	0.036	0.244	0.282	0.280	0.341

Notes: Ordinary least squares estimates. Dependent variable is indicated in the column header. Sample: respondents aged 16-65 years in the PIAAC survey who participated in PIAAC 2011/2012 and PIAAC-L 2014. *High-SES* indicates whether the respondent had more than 25 books at home at the age of 16. All behavioral traits are standardized to have mean zero and standard deviation one. All specifications control for a quadratic polynomial in age, gender, and migration status, as well as for imputation dummies for personality traits. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Data source: PIAAC 2011/2012, PIAAC-L 2014.

Table F3: Trust and Labor-Market Outcomes

	Unemployed		Monthly wages		Hourly wages	
	(1)	(2)	(3)	(4)	(5)	(6)
Trust	-0.022**	-0.018*	0.162***	0.155***	0.081***	0.072***
	(0.010)	*(0.010)*	*(0.034)*	*(0.034)*	*(0.022)*	*(0.020)*
Trust x Higher-SES	0.008	0.006	-0.037	-0.036	0.029	0.026
	(0.012)	*(0.011)*	*(0.039)*	*(0.039)*	*(0.024)*	*(0.023)*
Higher-SES	-0.007	-0.003	0.163***	0.132***	0.145***	0.114***
	(0.010)	*(0.010)*	*(0.038)*	*(0.038)*	*(0.024)*	*(0.024)*
Openness		0.001		-0.020		0.016
		(0.005)		*(0.017)*		*(0.011)*
Conscientiousness		-0.007		-0.017		-0.046***
		(0.005)		*(0.019)*		*(0.012)*
Extraversion		0.008**		-0.042***		-0.041***
		(0.004)		*(0.017)*		*(0.010)*
Agreeableness		0.012***		-0.075***		-0.046***
		(0.004)		*(0.017)*		*(0.010)*
Neuroticism		0.013***		0.003		-0.011
		(0.005)		*(0.018)*		*(0.011)*
Grit		-0.007		0.135***		0.064***
		(0.005)		*(0.020)*		*(0.012)*
Internal locus of control		0.004		-0.014		-0.006
		(0.004)		*(0.017)*		*(0.010)*
External locus of control		-0.009**		0.082***		0.070***
		(0.005)		*(0.018)*		*(0.011)*
Risk attitude		0.007		0.006		-0.027***
		(0.005)		*(0.017)*		*(0.010)*
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Imputation dummies for personality traits	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,019	3,019	2,637	2,637	2,419	2,419
R^2 (adjusted)	0.020	0.034	0.250	0.282	0.306	0.341

Notes: Ordinary least squares estimates. Dependent variable is indicated in the column header. Sample: respondents aged 16-65 years in the PIAAC survey who participated in PIAAC 2011/2012 and PIAAC-L 2014. *High-SES* indicates whether the respondent had more than 25 books at home at the age of 16. All behavioral traits are standardized to have mean zero and standard deviation one. All specifications control for a quadratic polynomial in age, gender, and migration status, as well as for imputation dummies for personality traits. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Data source: PIAAC 2011/2012, PIAAC-L 2014.

Table F4: Professional Qualifications by SES Background

	Low-SES	Higher-SES	Difference	Observations	
	Mean	Mean	*p*-value	Low-SES	Higher-SES
	(1)	(2)	(3)	(4)	(5)
No qualification					
Above age 25	0.21	0.06	*0.000*	979	3,223
Above age 30	0.21	0.06	*0.000*	907	2,772
Above age 35	0.20	0.05	*0.000*	818	2,404
Above age 40	0.18	0.05	*0.000*	717	2,013
Apprenticeship (> age 30)					
Dropout	0.08	0.05	*0.001*	900	2,702
Successful completion	0.71	0.67	*0.033*	907	2,772
University (> age 30)					
Dropout	0.02	0.06	*0.000*	900	2,702
Successful completion	0.06	0.26	*0.000*	907	2,772

Notes: Table shows group means by SES background. Respondents in the low-SES (higher-SES) sample had at most (more than) 25 books at home at the age of 16. Column 3 shows the *p*-value from a *t*-test comparing the mean of the respective variable across groups. Information on dropout is not available for individuals who were still enrolled in formal education at the time of the PIAAC interview. Statistics weighted by sampling weights. Data source: PIAAC 2011/2012.

Table F5: Professional Qualifications by SES Background: Individuals with Lower-Track Secondary Education

	Low-SES	Higher-SES	Difference	Observations	
	Mean	Mean	*p*-value	Low-SES	Higher-SES
	(1)	(2)	(3)	(4)	(5)
No qualification					
Above age 25	0.12	0.06	0.001	474	1,237
Above age 30	0.12	0.06	0.002	437	1,126
Above age 35	0.11	0.05	0.002	404	1,034
Above age 40	0.10	0.05	0.005	349	900
Apprenticeship (> age 30)					
Dropout	0.07	0.06	0.293	434	1,111
Successful completion	0.86	0.87	0.523	437	1,126
University (> age 30)					
Dropout	0.01	0.01	0.949	434	1,111
Successful completion	0.02	0.06	0.000	437	1,126

Notes: Table shows group means by SES background. Sample includes only individuals who obtained their highest school-leaving certificate from low-track (*Hauptschulen*) or intermediate-track (*Realschulen*) secondary schools. Respondents in the low-SES (higher-SES) sample had at most (more than) 25 books at home at the age of 16. Column 3 shows the *p*-value from a *t*-test comparing the mean of the respective variable across groups. Information on dropout is not available for individuals who were enrolled in formal education at the time of the PIAAC interview. Statistics weighted by sampling weights. Data source: PIAAC 2011/2012.

Table F6: Math Grades at School and Professional Qualifications

	No qualification		Apprenticeship completion		Apprenticeship dropout	
	(1)	(2)	(3)	(4)	(5)	(6)
Math grade	-0.025***	-0.023***	-0.015*	-0.012	-0.012**	-0.010*
	(0.005)	*(0.005)*	*(0.008)*	*(0.009)*	*(0.005)*	*(0.005)*
Math grade x Higher-SES	-0.002	0.000	0.001	0.001	0.000	-0.001
	(0.001)	*(0.001)*	*(0.001)*	*(0.001)*	*(0.001)*	*(0.001)*
Higher-SES	-0.049***	-0.023	-0.000	-0.009	-0.013	-0.024
	(0.015)	*(0.014)*	*(0.019)*	*(0.020)*	*(0.012)*	*(0.015)*
German grade		-0.008		0.001		-0.006
		(0.006)		*(0.009)*		*(0.006)*
German grade x Higher-SES		-0.002**		-0.002*		0.000
		(0.001)		*(0.001)*		*(0.000)*
Foreign-language grade		-0.007		-0.013		0.000
		(0.007)		*(0.010)*		*(0.006)*
Foreign-language grade x Higher-SES		-0.000		0.001*		0.000
		(0.000)		*(0.001)*		*(0.000)*
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
School-type fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Grade imputation dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,633	2,633	2,633	2,633	2,582	2,582
R^2 (adjusted)	0.168	0.184	0.324	0.332	0.041	0.047

Notes: Ordinary least squares estimates. Dependent variables: dummy for no professional qualification obtained (columns 1 and 2), dummy for completed apprenticeship (columns 3 and 4), and dummy for apprenticeship dropout. Grades are standardized to have mean zero and standard deviation one. *High-SES* indicates whether the respondent had more than 25 books at home at the age of 16. Sample: respondents aged 31-65 years in PIAAC who participated in PIAAC 2011/2012 and PIAAC-L 2014. Information on dropout is not available for individuals who were enrolled in formal education at the time of the PIAAC interview. All specifications control for a quadratic polynomial in age, gender, migration status, as well as for fixed effects for the type of secondary school that respondents attended and imputation dummies for school grades. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Data sources: PIAAC 2011/2012, PIAAC-L 2014.

Table F7: Math Grades at School and Professional Qualifications: Individuals with Lower-Track Secondary Education

	No qualification		Apprenticeship completion		Apprenticeship dropout	
	(1)	(2)	(3)	(4)	(5)	(6)
Math grade	-0.040***	-0.033***	0.020*	0.020*	-0.016**	-0.014*
	(0.007)	(0.007)	(0.010)	(0.010)	(0.008)	(0.008)
Math grade x Higher-SES	-0.001	-0.004*	0.000	0.004**	0.000	-0.000
	(0.001)	(0.002)	(0.000)	(0.002)	(0.000)	(0.001)
Higher-SES	-0.024	-0.011	-0.004	-0.017	-0.014	-0.024
	(0.015)	(0.016)	(0.019)	(0.020)	(0.015)	(0.017)
German grade		-0.011		0.001		-0.004
		(0.008)		(0.011)		(0.009)
German grade x Higher-SES		0.004*		-0.005***		0.000
		(0.002)		(0.002)		(0.001)
Foreign-language grade		-0.008		-0.014		-0.001
		(0.009)		(0.012)		(0.008)
Foreign-language grade x Higher-SES		-0.000		0.001		0.001
		(0.000)		(0.001)		(0.000)
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
School-type fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Grade imputation dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,563	1,563	1,563	1,563	1,545	1,545
R^2 (adjusted)	0.099	0.109	0.016	0.022	0.039	0.041

Notes: Ordinary least squares estimates. Dependent variables: dummy for no professional qualification obtained (columns 1 and 2), dummy for completed apprenticeship (columns 3 and 4), and dummy for apprenticeship dropout. Grades are standardized to have mean zero and standard deviation one. *High-SES* indicates whether the respondent had more than 25 books at home at the age of 16. Sample: respondents aged 31-65 years in PIAAC who participated in PIAAC 2011/2012 and PIAAC-L 2014 and who obtained their highest school-leaving certificate from low-track (*Hauptschulen*) or intermediate-track (*Realschulen*) secondary schools. Information on dropout is not available for individuals who were enrolled in formal education at the time of the PIAAC interview. All specifications control for a quadratic polynomial in age, gender, migration status, as well as for fixed effects for the type of secondary school that respondents attended and imputation dummies for school grades. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Data sources: PIAAC 2011/2012, PIAAC-L 2014.

Appendix G: Mediation Analysis

This appendix presents the setup of the mediation analysis (Appendix G.1) and its implementation for the low-SES (Appendix G.2) and higher-SES samples (Appendix G.3).

G.1 Setup of the Mediation Analysis

The mediation analysis follows the approach developed in Heckman, Pinto, and Savelyev (2013) and Heckman and Pinto (2015) (see also Oreopoulos, Brown, and Lavecchia (2017) for an application). Because of the opposing overall effects, we implement the mediation analysis separately for the samples of low-SES and higher-SES adolescents. Thus, our baseline equation 1 simplifies to a regression of the outcome Y_{it} on the treatment indicator T_i and baseline covariates $\boldsymbol{X}_{i(t-1)}$ in the respective subsample:

$$Y_{it} = \beta_0 + \beta_1 T_i + \boldsymbol{X}'_{i(t-1)}\boldsymbol{\beta_2} + \varphi_{it} \tag{G1}$$

To ensure that the mediation results are not affected by differences in the baseline covariates, we additionally control for the baseline values of the SES-specific mediator variables when available. As the sample split sometimes cuts through pairs with different SES, we do not use randomization-pair fixed effects in these specifications.[30]

The mediation approach assumes that the outcome can be expressed as a linear combination of k mediators M_{it}^k and a vector of baseline demographic characteristics $\boldsymbol{X}_{i(t-1)}$. This allows us to rewrite the previous equation as:

$$Y_{it} = \beta_0 + \beta_1^{residual} T_i + \sum_k \theta^k M_{it}^k + \boldsymbol{X}'_{i(t-1)}\boldsymbol{\beta_2} + \mu_{it} \tag{G2}$$

The setup implicitly assumes that any potential unobserved mediator subsumed in the error term μ_{it} is orthogonal to the included mediators and controls.

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 $1 - \beta_1^{residual}/\beta_1$.

[30] Appendix Table A9 shows that the subsample results are very similar with and without randomization-pair fixed effects.

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

$$M_{it}^{k} = \delta_0^k + \delta_1^k T_i + \boldsymbol{X}'_{i(t-1)}\boldsymbol{\delta_2^k} + \nu_{it} \quad \text{(G3)}$$

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:

$$share\ M_k = \theta^k \delta_1^k / \beta_1 \quad \text{(G4)}$$

In our empirical implementation, we combine the estimates of the different equations and calculate the explained and unexplained shares of the treatment effect by using the *nlcom* command in Stata.

G.2 Mediation Analysis for the Low-SES Sample

The primary aim of the mediation analysis is to investigate mediating factors of the significant positive treatment effect for low-SES adolescents. As potential mediators, we choose three variables, each elicited in both treatment and control group, that are related to facilitating low-SES adolescents' transition into professional life. By referring to schools, future orientation, and occupational orientation, the three mediators each relate to one of three components of our index of labor-market prospects. The first variable, *Perceive school as useful for job*, measures whether the respondent agrees with the statement that things learned in school could be useful for future jobs. It is measured on a 4-point scale and is standardized with a control-group mean of zero and a control-group standard deviation of one. The second variable, *Talk with mentor about future*, is a dummy variable that takes a value of one if the respondent talks to a mentor or coach about the future and zero otherwise. The third variable, *Mentor important for job choice*, is a dummy variable that takes a value of one if the respondent refers to information from a mentor or coach as being important for job choice and zero otherwise. Respondents could answer the survey question on the importance of the mentor or coach as a source of information for job choice on a 4-point scale from "very unimportant" to "very important," which we aggregate into a dummy variable taking a value of one if the mentor or coach is regarded as "rather important"

or “very important” and zero otherwise. The aggregation allows us to include the sizeable fraction of individuals (22 percent) who respond “I don’t know” in the non-important category.

The first columns of Tables G1-G3, respectively, show that for low-SES adolescents, the three mediator variables are significantly affected by the treatment in our baseline specification with the SES interaction (equation 1). Columns 1-3 of Table G4 confirm these results in the low-SES subsample (equation G3).[31]

Column 1 of Table G5 shows the overall treatment effect on the index of labor-market prospects for the low-SES sample (equation G1). Columns 2-5 show results when adding the mediator variables first individually and then jointly (equation G2). Comparing the treatment coefficients in the models with the individual mediators (columns 2-4) to the baseline model (column 1) yields the shares attributed to the three mediators that are depicted in the upper three bars of Panel A of Figure 3. When considered individually, *Perceive school as useful for job* accounts for 5 percent of the overall treatment effect for low-SES adolescents, *Talk with mentor about future* accounts for 29 percent, and *Mentor important for job choice* accounts for 17 percent.

The model that includes all three mediators jointly accounts for 34 percent of the overall treatment effect on the index of labor-market prospects for low-SES adolescents (comparison of columns 1 and 5: 1-0.427/0.644 = 0.34). Using equation G4 to assign shares to the individual mediators in the joint specification (shown in the fourth bar of Panel A of Figure 3), it becomes obvious that adolescents having a mentor as an attachment figure to talk about their future is by far the most relevant among the three mediators considered. In fact, the effect of the mentor being important for job choice materializes almost completely through talking with the mentor about the future.

Panel B of Figure 3 shows the mediation analysis for each of the three components of the index of labor-market prospects, based on columns 6-11 of Table G5. In the analysis, we follow Heckman, Pinto, and Savelyev (2013) and Oreopoulos, Brown, and Lavecchia (2017) in dropping mediators that would have a negative contribution in explaining the treatment effect, as the relative importance of the other mediators would be overestimated otherwise. The three mediators account for between 31 and 55 percent of the treatment effects on the three individual

[31] To be able to use the full sample in the mediation analysis, missing mediator values are imputed by the average in the treatment and control group, respectively, in Tables G4 and G5.

components. Surprisingly, the treatment effect of the mentoring program on math achievement in school is not mediated through perceiving school as useful for jobs, but rather through talking with the mentor about the future and conceiving the mentor important for job choice. As expected, the treatment effect on patience and social skills is primarily mediated through talking with the mentor about the future, and the treatment effect on labor-market orientation is primarily mediated through conceiving the mentor important for job choice.

G.3 Mediation Analysis for the Higher-SES Sample

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

From the mediators considered in the low-SES sample, *Talk with mentor about future* and *Mentor important for job choice* are also significantly positively affected by the treatment in the higher-SES sample (see bottom of first columns of Tables G2 and G3). However, since these variables also positively predict the index of labor-market prospects of higher-SES adolescents (albeit with very small effects; results not shown), they cannot explain the *negative* treatment effect on labor-market prospects in the higher-SES sample.

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

Results in column 1 of Appendix Table A14, column 2 of Table G1, and column 2 of Table G2 show that for higher-SES adolescents, these three mediators are negatively affected by the treatment. Columns 4-6 of Table G4 confirm these results in the higher-SES subsample, although the coefficient on talking with parents about the future becomes insignificant.

Comparing columns 12 and 13 of Table G5 indicates that the three mediators can account for 50 percent (=1-(-0.079/-0.156)) of the (small) negative treatment effect in the higher-SES sample. The decomposition analysis of equation G4 indicates that in the joint specification, the crowding-out of activities in school (41 percent) turns out to be by far the most relevant channel, while crowding-out of talking to parents about the future (8 percent) and reduced consideration of the importance of good school grades (2 percent) contribute less to the treatment effect.

Table G1: Effect of the Mentoring Program on Attitudes toward School and Parental Support Activities

	Attitudes toward school			Parental support activities	
	Perceive school as useful for job	Good grades are important	Recognition by teacher is important	Parental homework support	Received paid private teaching
	(1)	(2)	(3)	(4)	(5)
Treatment	0.359*	0.025	0.115	0.157	0.052
	(0.194)	*(0.229)*	*(0.190)*	*(0.171)*	*(0.067)*
Treatment x Higher-SES	-0.503*	-0.267	-0.269	-0.294	-0.051
	(0.291)	*(0.286)*	*(0.277)*	*(0.243)*	*(0.092)*
Higher-SES	0.302	0.057	0.090	0.371**	0.119
	(0.223)	*(0.280)*	*(0.235)*	*(0.185)*	*(0.077)*
Outcome in t_0	Yes	Yes	Yes	Yes	Yes
Randomization-pair fixed effects	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes
Observations	291	291	291	289	290
R^2	0.670	0.647	0.678	0.732	0.688
Treatment effect for Higher-SES	-0.144	-0.241*	-0.154	-0.137	0.001
	(0.166)	*(0.141)*	*(0.160)*	*(0.133)*	*(0.055)*

Notes: Table shows ITT effects of the mentoring program on attitudes toward school and support received from others. Dependent variable in column 1 measures on a 4-point scale whether the respondent agrees with the statement that things learned in school could be useful for a job. Dependent variables in columns 2 and 3 measure on a 5-point scale whether good grades or recognition by the teacher, respectively, is important. Dependent variable in column 4 measures on a 4-point scale to what extent parents provide support for homework. Variables in columns 1-4 are standardized. Dependent variable in column 5 is one if the individual received paid private teaching and zero otherwise. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Dummies for missing values in t_0 are included. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table G2: Effect of the Mentoring Program on which Attachment Figures Adolescents Talk to about their Future

	Mentor	Parents	Siblings	Other relatives	Friends	Teacher	Others
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.307***	-0.000	-0.016	0.052	-0.036	0.014	0.053
	(0.088)	*(0.067)*	*(0.091)*	*(0.104)*	*(0.101)*	*(0.102)*	*(0.057)*
Treatment x Higher-SES	0.174	-0.085	-0.010	-0.045	0.012	-0.038	-0.106
	(0.131)	*(0.092)*	*(0.140)*	*(0.154)*	*(0.135)*	*(0.166)*	*(0.093)*
Higher-SES	-0.127	0.111	0.146	0.229*	0.046	0.037	0.036
	(0.102)	*(0.078)*	*(0.106)*	*(0.132)*	*(0.101)*	*(0.127)*	*(0.092)*
Outcome in t_0	–	0.257***	0.517***	0.247**	0.268***	0.338***	0.073
		(0.080)	*(0.074)*	*(0.118)*	*(0.081)*	*(0.110)*	*(0.067)*
Randomization-pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	291	291	291	291	291	291	291
R^2	0.659	0.676	0.663	0.521	0.592	0.569	0.593
Treatment effect for Higher-SES	0.481***	-0.085*	-0.026	0.008	-0.025	-0.024	-0.053
	(0.075)	*(0.050)*	*(0.088)*	*(0.095)*	*(0.077)*	*(0.104)*	*(0.062)*

Notes: Table shows ITT effects of the mentoring program on the attachment figure to talk about the future. Dependent variable is one if the adolescent talks to the person indicated in the column header about the future and zero otherwise. Covariates are from the baseline survey and include: gender, age, received paid private teaching, parental homework support, and Big-5 personality traits. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table G3: Effect of the Mentoring Program on Importance of Different Sources of Information for Job Choice

	Mentor	Family	Friends	School	Employment agency	Media
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.261**	0.106	-0.011	0.040	0.031	0.048
	(0.101)	*(0.113)*	*(0.106)*	*(0.106)*	*(0.122)*	*(0.091)*
Treatment x Higher-SES	0.069	0.004	0.070	0.026	0.040	-0.121
	(0.156)	*(0.142)*	*(0.149)*	*(0.150)*	*(0.164)*	*(0.143)*
Higher-SES	0.012	0.008	-0.027	-0.055	0.029	-0.119
	(0.140)	*(0.120)*	*(0.117)*	*(0.129)*	*(0.125)*	*(0.129)*
Outcome in t_0	0.168	0.060	0.250**	0.264**	0.424***	0.255**
	(0.152)	*(0.103)*	*(0.097)*	*(0.116)*	*(0.100)*	*(0.099)*
Randomization-pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Observations	285	291	291	290	289	288
R^2	0.591	0.512	0.571	0.522	0.579	0.625
Treatment effect for Higher-SES	0.330***	0.110	0.058	0.066	0.071	-0.073
	(0.099)	*(0.069)*	*(0.088)*	*(0.097)*	*(0.085)*	*(0.098)*

Notes: Table shows ITT effects of the mentoring program on whether information from the entity indicated in the column header is important for job choice. Dependent variable is one if the indicated entity is rather important or very important for receiving information for job choice and zero otherwise. Covariates are from the baseline survey and include: gender, age, received paid private teaching, parental homework support, and Big-5 personality traits. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table G4: Effect of the Mentoring Program on Mediator Variables

	Low-SES sample			Higher-SES sample		
	Perceive school as useful for job	Talk with mentor about future	Mentor important for job choice	Activities in school	Good grades are important	Talk with parents about future
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.304*	0.309***	0.331***	-0.428***	-0.238*	-0.024
	(0.168)	*(0.070)*	*(0.077)*	*(0.149)*	*(0.132)*	*(0.044)*
Outcome in t_0	Yes	–	Yes	–	Yes	Yes
Randomization-pair fixed effects	No	No	No	No	No	No
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Observations	145	145	145	163	163	163
R^2	0.205	0.258	0.204	0.130	0.288	0.308

Notes: Table shows ITT effects of the mentoring program on mediator variables for the samples of low-SES and higher-SES adolescents, respectively. Dependent variable in column 1 measures on a 4-point scale whether the respondent agrees with the statement that things learned in school could be useful for a job. Dependent variables in columns 2 and 6 are one if the individual talks to the mentor or parents, respectively, about the future and zero otherwise. Dependent variable in column 3 is one if the mentor is important or very important for receiving information for job choice and zero otherwise. Dependent variable in column 4 is an average index of the following school activities (represented by a dummy variable that is one if true and zero otherwise): acting as class representative, working as peer mediator, acting as school representative, working for the school magazine, volunteering as school nurse, participating in the school music ensemble, participating in the school theater group, and participating in other school activity. Dependent variable in column 5 measures on a 5-point scale to what extent the adolescent considers good grades in school as important. Variables in columns 1-5 are standardized. Missing outcome values are imputed by treatment-specific averages. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Columns 1, 3, 5, and 6 additionally control for baseline values of the respective mediator variable. Dummies for missing values in t_0 are included. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table G5: Effect of the Mentoring Program on Labor-Market Prospects Conditional on Mediator Variables

	Low-SES											Higher-SES	
	Index of labor-market prospects					Math achievement		Patience and social skills		Labor-market orientation		Labor-market prospects	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Treatment	0.644***	0.609***	0.460***	0.535***	0.427***	0.217*	0.116	0.436***	0.199	0.465***	0.326*	-0.156	-0.079
	(0.150)	*(0.152)*	*(0.159)*	*(0.163)*	*(0.163)*	*(0.122)*	*(0.124)*	*(0.164)*	*(0.170)*	*(0.173)*	*(0.176)*	*(0.132)*	*(0.133)*
Perceive school as useful for job		0.110			0.084				0.171*		0.029		
		(0.073)			*(0.066)*				*(0.092)*		*(0.088)*		
Talk with mentor about future			0.598***		0.518**		0.260		0.601***				
			(0.193)		*(0.204)*		*(0.178)*		*(0.194)*				
Mentor important for job choice				0.329**	0.095		0.063				0.409**		
				(0.164)	*(0.164)*		*(0.155)*				*(0.167)*		
Activities in school													0.149**
													(0.059)
Good grades are important													0.012
													(0.064)
Talk with parents about future													0.504**
													(0.211)
Outcome in t_0	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Randomization-pair fixed effects	No	No	No	No	No	No	No	No	No	No	No	No	No
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	142	142	142	142	142	134	134	135	135	135	135	162	162
R^2	0.507	0.515	0.547	0.523	0.553	0.599	0.612	0.336	0.415	0.289	0.321	0.322	0.358

Notes: Table shows ITT effects of the mentoring program on labor-market prospects controlling for mediator variables in the samples of low-SES and higher-SES adolescents, respectively. Dependent variable in columns 1-5 and 12-13 is the index of labor-market prospects. Dependent variable in columns 6-11 is the respective component indicated in the column header. See notes of Table G4 for variable definitions. Covariates are from the baseline survey and include: gender, age, migrant, received paid private teaching, parental homework support, and Big-5 personality traits. Mediators are excluded if they would have a negative contribution in explaining the treatment effect. Columns 1-11 additionally control for baseline values of *Mentor important for job choice* and *Perceive school as useful for job*. Columns 12 and 13 additionally control for baseline values of *Good grades are important* and *Talk with parents about future*. Dummies for missing values in t_0 are included. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix H: Cost-Benefit Analysis

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

Present value of lifetime earnings. We use a representative dataset of German adults, PIAAC (see Appendix F), to calculate discounted lifetime earnings separately for low-SES adults (i.e., at most 25 books at home at the age of 16) and higher-SES adults (more than 25 books at home at the age of 16). We first calculate annual earnings by multiplying monthly earnings by 12, and express this value in 2017 Euros (as the program start was in 2017 for most of the adolescents in our sample). We smoothen the actual earnings stream by using predicted earnings from a regression of earnings on a quartic polynomial in age. We take into account that the age of labor-market entry differs by qualification (with 18 being the earliest entry age)[32] and assign persons before hypothetical labor-market entry zero earnings. We assume that persons exit the labor market at age 65.[33] Finally, to correct for periods of unemployment, we assign unemployed in our sample the standard rate of unemployment benefits. We discount future earnings at a net annual rate of 1.5 percent, which is comprised of a gross discount rate of 3 percent (e.g., Chetty et al. (2011); Heckman, Pinto, and Savelyev (2013); Lavecchia, Oreopoulos, and Brown (2020)) and a rate of potential output growth of 1.5 percent (Hanushek and Woessmann (2011); Hanushek, Ruhose, and Woessmann (2017)).

Although we find positive program effects on several outcomes, the cost-benefit analysis relies on the math-grade estimates. We do so for three reasons. First, as discussed in section 4.2, math achievement at school is highly predictive of future earnings. Second, among the outcomes studied in the paper, we deem math achievement as a measure of cognitive skills as most important for future labor-market success. Third, we know of no representative dataset that contains all variables necessary to construct our main outcome measure, the index of labor-

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

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

market prospects, and to assess the earnings benefits of an increase in this index. We use data on math grades from PIAAC, which elicits math grades from the end of formal schooling.

To take into account that the standard deviations in math grades differ between PIAAC and our mentee sample, our program effect estimates in the cost-benefit analysis use non-standardized math grades. Our baseline specification (see column 1 of Table 3) yields a significant program effect of 0.405 grade points for low-SES adolescents, while the program effect for higher-SES adolescents is not statistically significant. In PIAAC, we find that an increase in math grades by one grade point is associated with a monthly wage increase of 7.4 percent for low-SES individuals in the baseline model (equivalent to column 9 of Appendix Table F1). Multiplying the present value of lifetime earnings by the gain in monthly wages through better math grades and by the treatment effect on math grades, we estimate that the gain in discounted lifetime earnings from the program is about 23,500 EUR for low-SES participants.[34] Since program participation does not lead to significant grade effects for higher-SES adolescents, their earnings benefits are assumed to be zero. Weighting the benefits of low-SES and higher-SES participants by the sample share of the respective group, we arrive at overall earnings benefits of the program of about 11,000 EUR.

Program costs. According to the program's annual report, its total organizational costs amounted to 1,046,750 EUR in 2017.[35] Our best estimate of the number of mentoring pairs in operation in 2017 is about 1,400.[36] Thus, direct program costs are roughly 750 EUR per mentee.

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

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

[35] See https://rockyourlife.de/transparenz/.

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

Benefit-cost ratios. Table H1 reports benefit-cost ratios for different assumptions regarding (a) the discount rate, (b) program costs (with or without opportunity costs), and (c) program participants (with or without higher-SES adolescents). In all cases, program benefits exceed the costs to a sizeable extent. In our preferred specification with a net discount rate of 1.5 percent and no opportunity costs, the estimated program benefits outweigh costs by as much as 31-to-1 (23,500 EUR/750 EUR) if the program was targeted only at low-SES adolescents. If the program would not preselect only low-SES adolescents, the benefit-to-cost ratio would be 15-to-1 (11,000 EUR/750 EUR). When opportunity costs are also considered, the program yields benefit-cost ratios of 12-to-1 and 6-to-1, respectively. The large differences in the benefit-cost ratios by target group of the program indicates that the program foregoes substantial gains by not properly pre-screening participants. In fact, benefit-cost ratios would roughly double if the program would focus on the half of its subject pool that can be considered most disadvantaged.

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

Table H1: Benefit-Cost Ratios

	Actual costs		Actual costs and opportunity costs of voluntary work	
	Program targeted at low-SES	Untargeted program	Program targeted at low-SES	Untargeted program
Discount rate	(1)	(2)	(3)	(4)
0.0%	45-to-1	21-to-1	17-to-1	8-to-1
1.5%	31-to-1	15-to-1	12-to-1	6-to-1
3.0%	23-to-1	11-to-1	9-to-1	4-to-1

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

Appendix References

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

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.

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

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

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

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

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

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

Falck, Oliver, Alexandra Heimisch-Roecker, Simon Wiederhold (2020). Returns to ICT Skills. *Research Policy*: forthcoming.

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

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

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

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

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

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

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

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

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

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

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

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

Kristen, Cornelia, David Reimer, Irena Kogan (2008). Higher Education Entry of Turkish Immigrant Youth in Germany. *International Journal of Comparative Sociology* 49 (2-3): 127-151.

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

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

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

OECD (2013). *PISA 2012 Results: What Students Know and Can Do – Student Performance in Mathematics, Reading and Science (Volume I)*. Paris: Organisation for Economic Co-operation and Development.

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

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

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

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

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

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

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.

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

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